João Manuel de Oliveira Valente Morais

Licenciatura em Engenharia Electrotécnica e de Computadores

Unsupervised Classification of Uterine Contractions Recorded Using Electrohysterography

Dissertação para obtenção do Grau de Mestre em

Engenharia Eletrotécnica e de Computadores

Orientador: Arnaldo Batista, Assistant Professor,
Universidade Nova de Lisboa, FCT

Júri

Presidente: Doutor Rodolfo Alexandre Duarte Oliveira - FCT/UNL
Arguente: Doutor Raúl Eduardo Capelo Tello Rato - FCT/UNL
Vogal: Doutor Arnaldo Manuel Guimarães Batista - FCT/UNL

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março, 2019

Unsupervised Classification of Uterine Contractions Recorded Using Electrohysterography

Este documento foi gerado utilizando o processador (pdf)LaTeX, com base no template "novathesis" [1] desenvolvido no Dep. Informática da FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

Acknowledgements

I thank my thesis advisor Arnaldo Batista for the support and guidance in the course of this project. I'm grateful to Universidade Nova de Lisboa and it's associates who made part of an impressive experience that opened my horizons to many areas that I learned to love and become curious about.

A very special thanks to Paulo Oliveira for all the wise advice, patience and availability.

To my whole family and friends for all the great moments we share together.

To my partner and her parents for their friendship, unfailing support and continuous encouragement.

Lastly, I would like to express my most profound gratitude to my parents, brothers and grandmother for everything they have done for me that I cannot thank enough.

Abstract

Pregnancy still poses health risks that are not attended to by current clinical practice motorization procedures. Electrohysterography (EHG) record signals are analyzed in the course of this thesis as a contribution and effort to evaluate their suitability for pregnancy monitoring.

The presented work is a contributes with an unsupervised classification solution for uterine contractile segments to FCT's Uterine Explorer (UEX) project, which explores analysis procedures for EHG records.

In a first part, applied processing procedures are presented and a brief exploration of the best practices for these. The procedures include those to elevate the representation of uterine events relevant characteristics, ease further computation requirements, extraction of contractile segments and spectral estimation.

More detail is put into the study of which characteristics should be chosen to represent uterine events in the classification process and feature selection methods. To such end, it is presented the application of a principal component analysis (PCA) to three sets: interpolated contractile events, contractions power spectral densities, and to a number of computed features that attempt evidencing time, spectral and non-linear characteristics usually used in EHG related studies.

Subsequently, a wrapper model approach is presented as a mean to optimize the feature set through cyclically attempting the removal and re-addition of features based on clustering results. This approach takes advantage of the fact that one class is known beforehand to use its classification accuracy as the criteria that defines whether the modification made to the feature set was ominous.

Furthermore, this work also includes the implementation of a visualization tool that allows inspecting the effect of each processing procedure, the uterine events detected by different methods and clusters they were associated to by the final iteration of the wrapper model.

Keywords: Electrohysterogram (EHG), Electromyogram (EMG), Unsupervised Classification, Machine Learning, Signal Processing.

Resumo

A gravidez ainda comporta riscos para a mulher que não são, actualmente, contemplados pelos processos de monitorização praticados clinicamente. Esta tese analisa sinais de registos obtidos por electrohisterografia, como uma contribuição para a sua viabilidade na monitorização da gravidez.

Este trabalho tem seguimento do projeto *Uterine Explorer* (UEX) que acumula processos para análise de registos EHG. O trabalho aqui proposto contribui para o projeto com uma solução para a classificação não supervisionada de contrações uterinas.

Primeiramente, são apresentados os procedimentos aplicados para processamento dos sinais. Estes incluem processos que realçam as características relevantes dos eventos uterinos, baixam os requisitos de processamento, extraem os segmentos de contrações uterinas e estimam representações espectrais.

Mais detalhe é dado ao estudo de quais as características usadas para representar os eventos uterinos no processo de classificação e aos processos que analisam estas definir um subconjunto ótimo. Para este fim, é apresentada a aplicação da análise dos componentes principais (PCA) em três conjuntos de características distintos: eventos uterinos que são interpolados para possuírem todos a mesma dimensão, densidade espectral de potência estimada e um conjunto de características computadas que procuram evidenciar características temporais espectrais e de não-linearidade.

Posteriormente, é usada uma abordagem de modelo envolvente de agrupamento para otimização do conjunto de características a usar. Este, consiste na adição e remoção de características ao subconjunto que é usado num processo de agrupamento a cada iteração do modelo, com base nos resultados dos grupos criados. Esta abordagem tira proveito do facto de uma das classes pretendidas ser conhecida, pelo que a precisão na classificação desta é usada como critério para definir o impacto que a adição ou remoção de uma característica tem na solução. Desta forma o modelo envolvente providencia uma solução à classificação dos sinais, mas também otimiza o conjunto de características a usar.

Este trabalho inclui também a implementação de uma ferramenta de visualização que permite inspecionar o efeito dos diferentes processamentos feitos aos sinais, os eventos uterinos extraídos por diferentes métodos e os grupos a que estes ficaram associados na última iteração do modelo envolvente.

Palavras-chave: Electrohisterograma (EHG), Electromiograma (EMG), Classificação não supervisionada, Aprendizágem automática, Processamento de sinais.

Contents

List of Figures xiii

List of Tables xvii

Acronyms xix

1 Introduction 1
- 1.1 Context . . . 1
- 1.2 Motivation . . . 2
- 1.3 Objectives and Contributions . . . 3

2 State of the Art 5
- 2.1 Electrohysterogram . . . 5
- 2.2 Classification of Electrohysterogram Signals . . . 7
- 2.3 Uterine Explorer Tool . . . 8

3 EHG Signal Description and Preprocessing 11
- 3.1 Pregnancy Monitoring and Electrohysterography . . . 11
  - 3.1.1 Used EHG Records . . . 13
  - 3.1.2 Uterine Physiological Events . . . 14
- 3.2 Signal Preprocessing . . . 15
  - 3.2.1 Decimation . . . 16
  - 3.2.2 Filtering Process . . . 17
  - 3.2.3 Bipolar Signals . . . 19
  - 3.2.4 Extraction of Uterine Events . . . 21
  - 3.2.5 Signal Modeling . . . 25
  - 3.2.6 Spectral Analysis Methodologies . . . 26
  - 3.2.7 Spectral Analysis of Uterine Events . . . 29

4 Dimensionality Reduction 35
- 4.1 Feature Extraction . . . 35
  - 4.1.1 Introduction . . . 35
  - 4.1.2 Linear Time Domain Features . . . 36
  - 4.1.3 Spectral Features . . . 38
  - 4.1.4 Non-Linear Features . . . 40
  - 4.1.5 General Feature Extraction Methods . . . 41
- 4.2 Feature Selection . . . 43
  - 4.2.1 Filter Approaches . . . 44

4.2.2 Wrapper Approaches . . . 45
4.3 Defined Features . . . 46
4.3.1 Introduction . . . 46
4.3.2 Feature Sets . . . 47

5 Machine Learning **49**
5.1 Introduction . . . 49
5.2 Unsupervised Learning . . . 50
5.2.1 Data Clustering . . . 50
5.2.2 Classification Methods used in Related Studies . . . 54
5.3 Validation . . . 55
5.4 Classification Model For EHG Data . . . 55
5.4.1 PCA and K-means Combination . . . 56
5.4.2 Wrapped Clustering Model . . . 57

6 Project Implementation **61**
6.1 Introduction . . . 61
6.2 Signal Processing and Classification Program . . . 61
6.2.1 Signal Processing . . . 61
6.2.2 Feature Selection and Classification Procedures . . . 67
6.3 Visualization Tool . . . 71
6.4 Results . . . 74
6.4.1 Principal Component Analysis . . . 75
6.4.2 Wrapped Clustering Algorithm . . . 78
6.4.3 Selected Subset Features . . . 81

7 Conclusions **87**
7.1 Future Work . . . 88

Bibliography **91**

List of Figures

2.1 Detection and classification in uterine EMG. The numbers on top refer to: '1' - background activity; '2' and '6' - contractions; '4' and '7' - fetus motions; '5' - LDBF waves; '3' - Alvarez waves. [11] 6
2.2 First illustrated example in the literature of the different uterine events [12] . 7

3.1 Typical example of the 4x4 electrodes matrix and TOCO sensor positioned on the woman's abdomen [19]. 13
3.2 EHG Electrodes Placement [3] 14
3.3 EHG Raw Signal. The amplitude has no meaning as it is relative due to machine gain. 16
3.4 Effect of decimation. The plots represent a signal before and after applying the decimation filter, respectively. 17
3.5 Effect of applying a pass-band filter with cutoff frequencies in 0.1 and 1Hz. The plots represent the signal before and after applying the filter, respectively. . . 19
3.6 Relation of monopolar to bipolar signals. The top-most and left-most circles represent the calculated bipolar signals, which specify their respective numeration and the computed subtraction. 20
3.7 Effect of subtracting monopolar signals. The first two graphs represent records from different electrodes, while the third represents the bipolar signal resultant of the subtraction. 21
3.8 Detected contractions using Wavelet energy 23
3.9 Detected contractions using Teager-Kaiser energy 24
3.10 Detected contractions using RMS 24
3.11 Detected contractions using RMS squared 24
3.12 Detected contractions using Hilbert-Huang spectrum energy 24
3.13 PSD representation of a contraction by different methods 30
3.14 Frequency Response of Hanning, Hamming and Blackman Windows in decibels 32
3.15 Frequency Response of Hanning, Hamming and Blackman Windows in decibels 32

4.1 Determination of number of sources by eigenvalue ordering [64, pp. 182] . . . 43
4.2 Flow diagram of wrapper approach for unsupervised learning [67, pp. 848] . . 45

5.1 Example of how normalization may reduce cluster separation [69] 51
5.2 Developed Wrapper Model for feature selection. Green and blue areas refer to SBS and SFS parts, respectively. The diamond boxes represent conditions, whereas rectangular boxes represent stages or procedures. 58

6.1 Flow diagram summarizing signal processing procedures. Pink boxes refer to functions external to the developed program, which belong to UEX project scripts. Green boxes refer to the scripts developed that implement the main flow of the program. Scripts names are are shown with a cyan background. The beige boxes represent data. The diagram provides a legend for arrows in the grey box. 62

6.2 Example output of script execution. Provides a structure with patient case profile and processed signal. 63

6.3 Example output of a1_2_set_Classif_WS.m script execution. Provides a structure with patient case profile and uterine event extracted. 64

6.4 Example output of a2_set_Classif_WS.m script execution. Provides a structure with each contraction extracted per row and information of where it was extracted from along with respective patient case profile. 64

6.5 Example output of a3_PWDWS.m script execution. Provides a structure with each contraction's spectral representations by different methods computed in a UEX function execution. 65

6.6 Example output of a4_FeatCalc.m script execution. Aggregates computed values that attempt to elevate the uterine event's time, spectral and non-linear characteristics. Each row of computed values is associated to an uterine event. 66

6.7 Flow diagram summarizing the developed procedures for feature selection and uterine events classification. Green boxes refer to the scripts developed that implement the main flow of the program. Script names are shown with a cyan background. The beige boxes represent data. The diagram provides a legend for arrows in the grey box. 67

6.8 Example output of a6_computePCA.m script execution. This output exemplifies a file generated when applying PCA to the uterine event in the time domain. The first 10 rows provide identification and profile information, and in specific the first row specifies whether the row corresponds to a fetal movement. Column 12 has the contractile events considered in the PCA procedure. These all present the same size because each uterine event was interpolated to be represented in a 512 sample signal. Column 13 shows the resulting transformed and reduced result of PCA execution. According to the evolution of the eigen vector rating, the transformed signals were reduced to only account for 20 samples. The last column shows the cluster assigned to each signal. 68

6.9 Output of wrapper model execution with information gathered over each of it's clustering iterations. 69

6.10 Table outputted from Wrapper model execution showing final clustering result of each uterine event. 70

6.11 Preprocessing part of visualization tool. 71

6.12 Preprocessing part of visualization tool. The drop-down list contains the preprocessing procedures that can be visualized, containing: 1-divide machine gain; 2-remove points from beginning and ending of signal, due to the amount of corrupted signals from moving electrodes when the test is starting and ending; 3-compute bipolar signals; 4-Remove signal tilt; 5-center signal to have zero average (this step was omitted from initial preprocessing description as it is not necessary); 6-subtract first value to the whole signal, for it to start at zero; 7-Result of Decimation; 8-Result of filtering to the band of interest. . . 72
6.13 *Signal Inspector* Menus. 72
6.14 Visualization Tool - *Signal Inspector*: Here the scope was set to 70 times the signal length (in *Signal Indexer* menu). in *Methods* menu three contraction detection methods are chosen, namely, by wavelet energy (red segments), Teager energy operator (black segments) and Hilbert-Huang (yellow segments). As such in the defined scope all contraction detected by these methods will appear with different colors. The small orange segment is the selected one. 73
6.15 Spectral representation of the selected contraction segment. Here, only the PSD representations were loaded, and the figure shows it's estimation according to Welch and autoregressive (AR) estimators. Although many methods of contraction detection are selected, the contraction segments iterated are those with respect to the first oneto be selected (subsequent are selected by holding CTRL key). 73
6.16 Cumulative Sum of the power representation. This plot simply integrates the estimated spectral representation to allow a visual inspection of how the energy evolves in the signal. 74
6.17 *SignalInspector* as a way to compare contraction detection methods. The orange signal segment a contraction detected using wavelet energy and the black segment represents a segment detected using Teager energy operator. . 75
6.18 Distribution of the interpolated and transformed time-domain fetal movements and contractions among the clusters. 76
6.19 Distribution of the transformed PSD of fetal movements and contractions among the clusters. 77
6.20 Distribution of the transformed computed features of fetal movements and contractions among the clusters. 77
6.21 Wrapper model applied to synthetic data. Each line represents the percentage of correctly classified signals and the features being added and removed are being represented with a + or −, respectively. 79
6.22 Iterations information of wrapper model. Blue line refers to the percentage of fetal movements being correctly classified, while the red line refers to the percentage of contractions being classified as fetal movements. 80
6.23 Distribution among clusters of the computed features selected by wrapper model and transformed by PCA algorithm. 82

6.24 Stacked Graph representing percentage of each clustered class associated to each gestational week. The weeks are not rounded up, meaning a new week only ends at day one of posterior week. Text at the top of each bars indicates number of uterine events (U.E.) and patient cases considered for that week. . 83
6.25 Stacked Graph representing percentage of each clustered class associated to each delivery week. The weeks are not rounded up, meaning a new week only ends at day one of posterior week. Text at the top of each bars indicates number of uterine events (U.E.) and patient cases considered for that week. 84
6.26 Stacked Graph representing number of contractions at each clustered class associated to each gestational week and delivery week. 85

List of Tables

3.1 Characteristics of commonly used windows [60] 33

4.1 Linear Time-Domain Features 38
4.2 Spectral Representations Considered in EHG Studies and Methods used . . . 38
4.3 Non-Linear Time-Domain Features 39
4.4 Non-Linear Time-Domain Features 41

5.1 General Minkowski Metric Derivations 51
5.2 Cluster Criteria Functions 54

Acronyms

ADASYN	Adaptive Synthetic Sampling.
AP	Action Potentials.
AR	Autoregressive.
BPSO	Binary Particle Swarm Optimization.
DFT	Discrete Fourier Transform.
ECG	Electrocardiography.
EEG	Electroencephalography.
EHG	Electrohysterography.
EMD	Empirical Mode Decomposition.
EMG	Electromyography.
FCBF	Fast Correlated Based Filter.
FFT	Fast-Fourier Transform.
FIR	Finite Impulse Response.
FNN	False Nearest Neighbor.
FT	Fourier Transform.
GS	Gramschmidt.
HHT	Hilbert Huang Transform.
IAP	Intracell Action Potentials.
ICA	Intrinsic Component Analysis.
IIR	Infinite Impulse Response.
IMF	Intrinsic Mode Functions.
IUPD	Intrauterine Pressure Catheter.
LAHF	Low Amplitude High Frequency.
LDA	Linear Discriminant Analysis.
LDAb	Linear discriminant analysis using backward search.

LDAf	Linear discriminant analysis using forward search.
LDAi	Linear discriminant analysis using independent search.
LDB	Local Discriminate Basis.
LDBF	Long Duration Low Frequency.
LS	Linear Spectrum.
LSD	Linear Spectral Density.
ML	Maximum Likelihood.
PCA	Principal Component Analysis.
PS	Power Spectrum.
PSD	Power Spectral Density.
RMS	Root Mean Square.
SBFS	Sequential Backward Feature Selection.
SBS	Sequential Backward Search.
SFFS	Sequential Forward Feature Selection.
SFS	Sequential Forward Search.
SMOTE	Synthetic Minority Over-sampling Technique.
SSE	Sum-of-squared-error.
STFT	Short-time Fourier Transform.
TF	Time-Frequency.
TFD	Time-Frequency Distributions.
TOCO	tocodynamometry.
uEMG	Uterine Electromyography.
WGN	White Gaussian Noise.
WPD	Wavelet Packet Decomposition.
WT	Wavelet Transform.
WVD	Wigner-Ville Distribution.

CHAPTER 1

Introduction

1.1 Context

Pregnancy holds many dangers both to the mother and fetus, being its monitoring a highly important prevention to diagnose any issues timely and take more successful interventions. Pregnancy monitoring covers a number of methods used to identify issues related the patient and fetus well-being.

More specifically, this thesis is concerned about assessment of uterine contractions as a way of pregnancy monitoring. The most widely used technique to assess contractions is the external tocodynamometry (TOCO). This is a harmless technique, which performs a mechanical measurement of the contractions based on pressure. It provides accurate information about the frequency and duration of contractions, but not their amplitude [1].

A more recently available method for uterine contractile activity assessment is electrohysterography (EHG). This is a noninvasive technique that measures action potentials in the uterine muscles through electrodes placed on the patients abdomen. EHG intent is to record the the coordinated adhesion of myometrial cells to propagating electrical action potentials, which forms the uterine contractile activity [1]. In an article by W. Cohen [1] it is claimed EHG is at least as reliable as TOCO.

While not yet well recognized and practiced, EHG is increasingly attaining better results on providing insightful information of uterine activity, which mat lead it to become a candidate for use in diagnosing pregnancy complications, such as to predict preterm labor or identifying fetal distress.

Detection of preterm labor is one of the major topics of studies regarding EHG records. Preterm labor is defined to be when it happens before the 37_{th} completed weeks and may bring complications including: neurological, mental, behavourial and pulmonary problems to the child [2].

In this work, the proposed approach envisions the use of detected contractions in EHG records as inputs of a classifier in a search of recognizing patterns. The EHG records that will be used are from a public database at PhysioNet page [3]. The intended result on this thesis is correctly differentiating four known types of uterine activity, namely: long amplitude high frequecy (LAHF) waves, Braxton-Hicks, long duration low frequency

(LDBF) waves and Fetal Movements.

This thesis takes part on a broader project: the UterineEXplorer (UEX) [4]. UEX is a tool developed for academic purposes at Universidade Nova de Lisboa that analyzes EHG signals. UEX contains some functions that will be used in this thesis, of which one performs the automatic detection of uterine events. This function has separate procedures to detect contractions and fetal movements, which makes this the only known class à priori.

The classification of events detected will, thus, be unsupervised meaning that the classifier will have no prior knowledge of what each event represents (other than fetal movements), and will make an attempt at identifying relevant patterns that provide an accurate identification of each contraction type.

Unsupervised classification or clustering is the process of finding groups that share similarities within them and dissimilarities between them. In this regard it is important to make relevant these characteristics that can distinguish these groups to obtain good results.

As such, it is requires the application of signal preprocessing techniques to improve the signal to noise ratio, make signals interpretable and extracting relevant information from them, followed by a classification procedure.

Ultimately, an affiliation between obtained classes and real types of contractions would determine the success of the study and provide functionality to the approach.

1.2 Motivation

Pregnancy monitoring is a preventive approach to problems such as abortion, premature-labor, fetus malformations and detection of any complications in the overall pregnancy and health of both the fetus and mother.

A computer-based approach for the analysis of EHG records in order to study pregnancy and it's monitoring methods provides a way to sustain more data into consideration along with more analysis complexity. This, can help determine the different contraction's characteristics and their differentiation along allowing the use of new variables to the decisions in pregnancy monitoring.

In this work it is intended the development of an unsupervised learning model to differentiate contractions extracted from EHG signals. Although the classes are not known à priori, which gives us no guarantees of how accurate or relevant the final classes became, the author believes it may be a step towards finding relevant patterns on the study of these signals. As such, the model developed should take into consideration how little it's known of these signals and how this may change, requiring the ability to evolve onto new tests and changes in our view of these signals.

At a current clinical practice level, it's mostly considered contractions frequency, amplitude and duration. However, through computer-based approaches to analyze EHG records, it may be possible to study pregnancy and it's monitoring methods considering other features of these signals or more complex phenomena. Computer-based approaches also

allow considering a greater amount of sample signals than that which is humanly possible, allowing more statistically relevant conclusions.

EHG is being studied as a non-invasive alternative that may lead to better results than commonly used methods clinically such as TOCO [5]. Successfully identifying types of contractions in an automatic way, may lead to further studying their meanings and identifying malicious patterns that can be used to diagnose pregnancy complications.

On the whole, EHG study is a promising approach and classifying contractions extracted from it may lead to find relevant patterns for it's study and ultimately it would be a great result if some of the classes obtained were to accurately represent real types of already identified contractions.

1.3 Objectives and Contributions

The objective of this thesis is to differentiate contractions extracted of EHG signals through an unsupervised classification model. Through this approach, the contribution to pregnancy monitoring field should be an analysis over contractile patterns and relevancy of features extracted from them. An ultimate goal would be to provide means of identifying what characterizes the different types of contractions and malicious patterns to pregnancy.

For the latter purpose, it should be designed and implemented an unsupervised learning model that allows testing features and a classification procedure to provide means of performing pattern study over EHG signals, as well as proposing a possible solution.

The contractions should be separated in a way able to identify different types of contractions on an automated way. The approach consists on building an unsupervised classifier trained by features extracted from EHG signals to define a set of classes that should be representative of the identified and documented real contractions.

It's ambitioned a successful choice of the best characteristics to define these kinds of signals and a learning method that weights well their distributions and nature.

Automatizing contraction categorization should be a step to better understand different kinds of contractions and to allow future work on identifying ominous patterns.

It is not assured in an unsupervised learning, however, that the obtained classes are relevant or, at least, it should be hard to ground the hypothesis. As such, it should also be implemented a way to ease inspection over the results to allow for it's visual validation by specialists, which is out of the scope of this thesis.

This work is also considering contractions that are automatically extracted from the EHG signals using methods from earlier works for the Uterine Explorer (UEX) project [4]. As such, the choice over what represents a contraction was already made and is not covered in this work.

Summing up, it should be proposed a solution to classify uterine events and test out their characteristics. This should provide a contribution towards identifying patterns in these signals and a way to easily inspect results through visual validation.

CHAPTER 2

State of the Art

2.1 Electrohysterogram

The study of uterine electrical phenomena dates from as early as 1910, when Theilhaber detected a significant difference in the action of the string galvanometer between the ovaries and uterus of non-menopause and menopause women [6].

Study of uterine contractile activity was usually performed with mechanical measurement of the contractions by either a tocotransducer, which is externally positioned on the abdominal wall with an elastic strap or by an intrauterine pressure (IUP) measurement [7].

An approach that measures uterine electrical activity, the electrohysterogram (EHG), was first recorded on humans using abdominal electrodes as early as 1931 [7]. The records were described as containing a 'slow' electrical wave in a band of 0.03-0.1Hz with an amplitude of 1-5mV, with an overlapping 'fast' wave in a band of 0.3-2Hz with an amplitude of 50μV-1mV [7].

Alvarez and Caldeyro [8] published in 1950 a study on uterine contractile activity of 120 clinically normal cases, summing 90 labors and 30 pregnancies, using an intrauterine pressure catheter and reported the first description of the low amplitude high frequency (LAHF) wave, also denominated Alvarez wave. The study described the existence of two waves and their evolutions along pregnancy: LAHF, which had never been reported and Braxton-Hicks, named after John Braxton Hicks, who first described these contractions in 1872.

A study on the significance of Alvarez waves by Newman et al. [1] hypothesized the possibility of Alvarez waves being more frequent in patients who subsequently developed preterm labor, although the statistical value made the theory inconclusive. This hypothesis was also statistically verified in the results from a study performed by W. Roberts et. al [9]. However, other studies state the existence of no correlation between the occurrence of LAHF contractions and parity nor gestational age and the difficulty of taking any conclusion given that these account for 70 to 80% of the total contractions recorded in normal pregnancies [10].

Marque et al. [7] studied external EHG records from different pregnancy stages, summing 11 labor, 6 pregnancy and 4 supplementary cases, corresponding to two induced labors, one premature labor risk and one labor with uncoordinated contractions. Results showed pregnancy contractile activity is mainly characterized by low frequencies and a long duration (averaging 74.6s), whereas that of labor is related to the presence of higher frequencies and shorter duration (averaging 59.3s).

M. Khalil et. al [11] performed the first contraction segmentation algorithm detecting four types of already identified uterine events: Alvarez waves, Long duration low frequency (LDBF) waves, contractions and fetal movements. In their work, they performed the decomposition of the signals using wavelet transform (WT) to obtain scales. These were, then, applied to an unsupervised classification process based on the on-line comparison of variance covariance matrices, which classified the events. Finally, a supervised algorithm based on neural networks was applied to identify each group of events. The result of the event detection and grouping algorithms are represented in Figure 2.1.

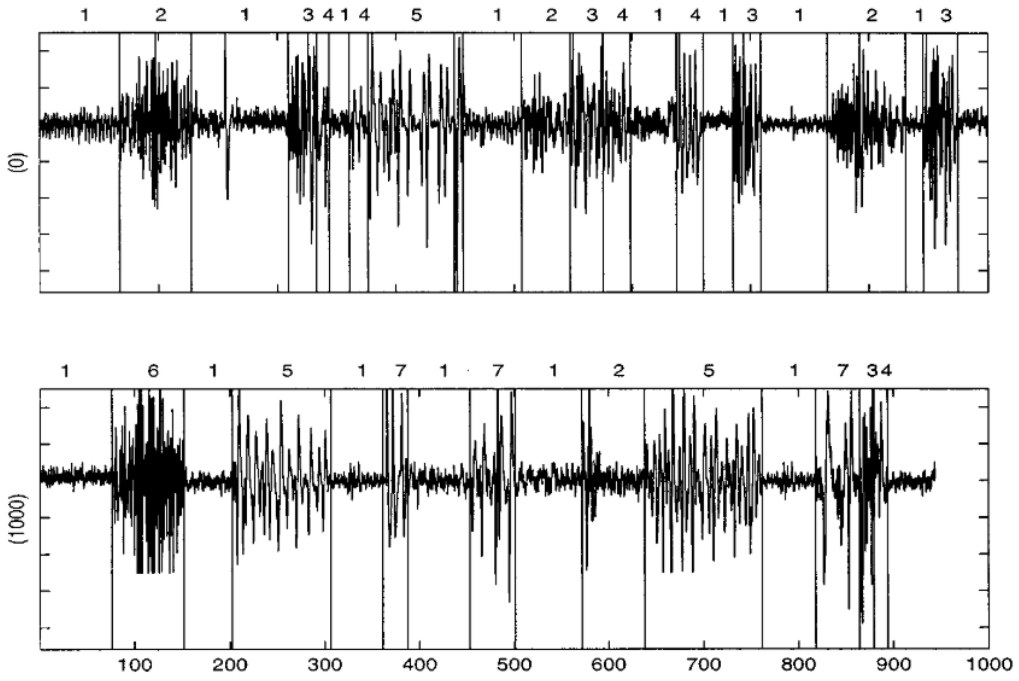
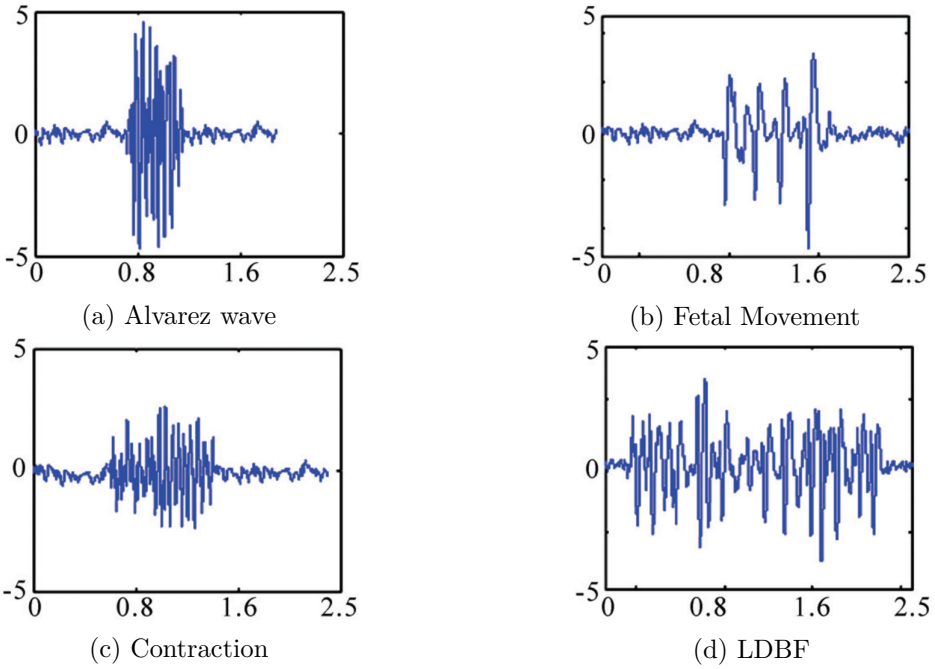


Figure 2.1: Detection and classification in uterine EMG. The numbers on top refer to: '1' - background activity; '2' and '6' - contractions; '4' and '7' - fetus motions; '5' - LDBF waves; '3' - Alvarez waves. [11]

Regarding the comparison of methods for uterine contraction activity, J. Alberola-Rubio et. al [5] make a comparison of EHG signals to the tocography alternative. The results defend that EHG-based techniques were able to detect a higher number of uterine contractions than tocography, especially from recordings taken over the uterine median axis. In another study, however, K. Horoba et.al [13] describe EHG signals as difficult to interpret, requiring computer-aided systems, and tocography as an inaccurate mechanical technique. The study disregards both for considering neither techniques provide an



(a) Alvarez wave (b) Fetal Movement

(c) Contraction (d) LDBF

Figure 2.2: First illustrated example in the literature of the different uterine events [12]

acceptable quantitative description of the recognized patters.

2.2 Classification of Electrohysterogram Signals

Classification of signals, refers to the application of procedures that differentiate the signals into classes. In the context of this thesis, these procedures have no knowledge of what the classes should look like, having to define them according to what they learn from the data that is provided. Hence, these procedures fall in a class of algorithms referred to as unsupervised learning.

Although it's not our goal it should be referred that some studies applied supervised learning algorithms, meaning it was used labeled or classified data in the learning procedure such as by delivery week [2, 14–17], or by whether it was a pregnancy or a labor signal [18].

In an unsupervised learning problem a crucial step is knowing the domain in which our classes play out to define how to influence the classifier to cluster them together and differentiate them from the other classes.

Studies regarding EHG signals have attempted to define the relevance of their different characteristics to provide better characterizations of their behaviors. These attempt at improving the significance of the information extracted from EHG signals to be used on different applications. In the literature, one can find descriptions of characteristics including:

- Signal propagation [19–21];

- Non-linearity measures [19, 21–24];
- Time-Frequency representations [4, 25, 26];
- Overall feature selection methods [15, 27]

Most found studies regarding EHG provide attempts at detecting preterm birth [12, 24, 28–31], contractions [11, 32, 33], differentiate pregnancy stages and labor [18, 19, 34, 35].

The literature refers to the application of unsupervised classification using EHG records, is studies including:

- M. Diab et. al [28] performed a Fisher test based on the comparison of the variance vectors computed from the scales, obtained from a Wavelet packet decomposition. The study considered contractions having the same registration week of delivery but different birth week of gestation (BWG) as an attempt to predict preterm. Results showed the classification error became high when the difference between BWG was small and small for big BWG divergence. As such, they obtained good results for cases where the distance between classes was large.
- M. Khalil et. al [12] othar than in [11] performed another unsupervised classification of contractions to difference contractions, foetus motions, Alvarez waves, LDBF waves using Support vector machines (SVM) using variances of the selected packets from a wavelet packet decomposition as features. The same features are used in another study by the same authors [36] which perform a classification based on Fisher's test and clustering to detect term and preterm delivery.

This work will attempt a classification procedure that clusters similar contractions together. An optimal result would differentiate real types of contractions such as Alvarez waves, Braxton-Hicks contractions, LDBF and fetal movements. However, the validation of whether the generated clusters really represent those would have to be made by a specialist which could visually correlate those contractions of a specific cluster with real ones.

2.3 Uterine Explorer Tool

Previous to classifying contractions, EHG records require preprocessing procedures mostly to enhance the influence of uterine events rather than unused band or noise from other sources and to evidence specific characteristics of the signals that may more accurately represent the types of contractions. Furthermore, it is also necessary to extract the contractions out of the original EHG signals, which represent full test recordings containing uneventful segments.

These procedures are mostly assured by earlier work under the Uterine Explorer (UEX) project [4] developed in MATLAB to analyze EHG signals. Currently, UEX offers means to perform different preprocessing procedures to the signal as well as analyze many of

it's features in time, spectral and time-frequency domains. It has the ability to perform contractile event extraction from the full signals, using prominence of peaks over signal energy as criteria to define what is a relevant event and it's boundaries.

UEX is referred here as the goal of this thesis is held in the context of bringing an unsupervised classification of contractions that may integrate the UEX tool project or contribute to a solution for that problem. Thus, other UEX contributing studies will have an influence over decisions to come and provide some of the procedures used.

CHAPTER 3

EHG Signal Description and Preprocessing

3.1 Pregnancy Monitoring and Electrohysterography

Pregnancy monitoring is a very important matter as it aims maintaining both the patient and fetus health [28], approaching problems such as miscarriage and premature labor. A number of methods are used for monitor pregnancy, with different applications and purposes, from assessing cardiac response and functioning to fetal development and other physiological events to identify anomalous scenarios. Pregnancy monitoring methods include, among others, ultrasonography, electrocardiography (ECG), phonography, phonocardiography and electrohysterography (EHG).

Ultrasonography consists on acquiring images from the selective reflexion of acoustic energy, which contrast tissues with different reflexion properties. This method can be used to verify if the fetus structures are well developed and even get motion characteristics of these structures. As an example, Cardiotocography is a type of ultrasonography that measures heart rate using a Doppler-ultrasound based system. While this method is very powerful there are concerns about continuous exposure to the ultrasonic signal.

A very commonly used method is Electrocardiography (ECG). It records fetal cardiac activity using electrodes that measure the electrical activity of the heart. This method can be performed indirectly by placing electrodes over the patients womb or in a invasive way when an electrode is placed directly on the fetus scalp. The former can be used during the pregnancy to diagnose congenital heart diseases or study the fetus heart rate response to drugs administered. The later approach is highly invasive and only possible during labor and under specific dilation conditions. This method may be used to assess fetal distress, which may be an indicator of unsuccessful contractions that should be addressed by adjusting oxitocin administration.

Phonography is a method used to measure vibrations on the patients abdominal wall through a transducer. It has evolved a lot over the years to become more sensible to fetal activity and to be able to record it even when the fetus changes position. Being a non-invasive method and not emitting energy to the fetus makes phonography a viable

candidate for continuous monitoring of fetal health during pregnancy.

Electrohysterography (EHG) or uterine electromyography (uEMG) measures the intracellular action potentials (IAP) of the uterine muscle cells through electrodes placed on the abdominal wall. Uterine electrical activity results from the depolarization and repolarization of myometrial smooth muscle cells [37] and is the source of EHG signals. The action potentials (APs) measured constitute neural stimulation for contraction and relaxation of cells and their propagation through the myometrium is an electrical burst recorded by the EHG electrodes [38]. These records are representative of the physiological events taking place in the uterus containing information about contractions and fetal movements [38]. The evolution of contractions throughout pregnancy is related to an increase in cellular excitability and synchronization, having more cells adhering to the contractions and is a necessary condition for enabling cervix dilation and fetus expulsion [15].

Interpretation of EHG records, even though has still not met the accuracy standards for application in clinical practice [18, 39], is increasingly getting attention from various studies. Researchers aim at applying the study of EHG records as a clinical practice for overall pregnancy monitoring and specific cases like contractions affectability on EHG records [33], contraction evolution and specific applications such as, mainly, detection of preterm labor. The later application is particularly important as it may carry neurological, mental, behavioral and pulmonary implications to the fetus and may even lead to death [40], along other social and economic implications [2, 14]. An accurate preterm delivery prediction method would allow better prevention, treatment and resource allocation [40].

Studies comparing EHG with the most commonly used clinical practices including intrauterine pressure determination (IUPD) and external tocodynamometry (TOCO) have shown that EHG is more reliable than TOCO, according to IUPD data, described as the golden standard, yet with the advantage of being non-invasive in comparison with IUPD [32, 41]. An article by L. Chen and Y. Hao [17] describes EHG method as a noninvasive, low-cost, real-time, and effective technique compared to traditional methods.

Other than unmet accuracy standards, EHG's applicability still meets some challenges as its records are complex and difficult to interpret, implying high dependence of a computer-aided system [13]. Not being a clinical practice, brings difficulties to collect a more statistically relevant amount of data that could further model these problems domain. Thus, an important step towards making EHG feasible may be creating standards for the acquisition of EHG records, such as for the number of electrodes and their position, as well as acquiring more data to increase statistical significance.

As previously mentioned EHG records are obtained through electrodes placed on the abdominal wall. However, there are still no standards for the application of this method and different setups for the data acquisition method are found among EHG studies. Within these, electrodes are usually placed 2.5 to 7cm horizontal and vertically apart [42], with slight changes on the overall position of the electrodes over the abdominal wall, different dispositions and number of electrodes. The later diverge between 2 [34], 4 [14, 24, 39, 43–45], 5 [46], 10 [47, 48], 16 [15, 17, 18, 23, 35] up to 64 [38]. An usual setup of a 16

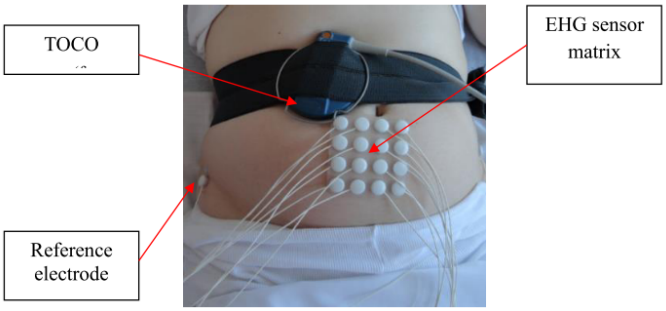


Figure 3.1: Typical example of the 4x4 electrodes matrix and TOCO sensor positioned on the woman's abdomen [19].

electrode EHG recording test and the simultaneous recording with a TOCO is illustrated in figure 3.1.

An increased number of electrodes becomes relevant considering the functional and structural complexity of the uterus [25]. This complexity arises from the stochastic, nonlinear mechanisms interacting with a fluctuating environment, which implies limitations in the use of a reduced number of sensors [25]. The use of multiple electrodes also allows spatial perception required for the study of propagation of the electrical activity throughout the uterus [19] and is required to attain bipolar signals which result from the subtraction of the record of an electrode with another to cancel common noise.

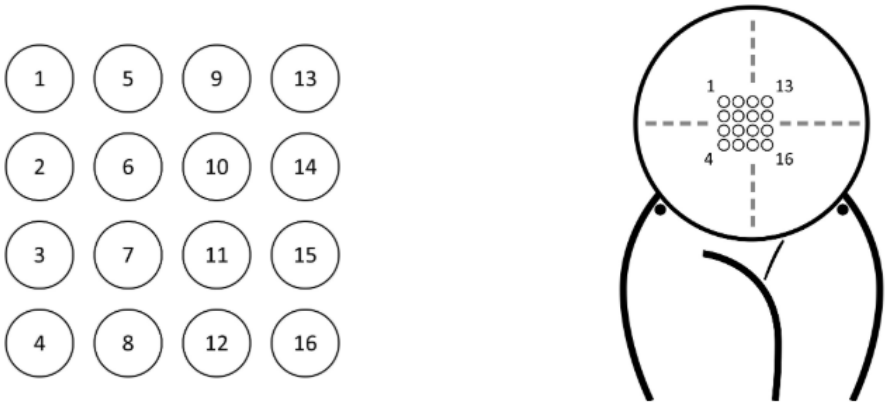
3.1.1 Used EHG Records

In this work it will be used EHG records from the *Icelandic 16-electrode Electrohysterogram (EHG) Database* stored in PhysioNet page [3]. The recording system consists of 16 electrodes arranged in a 4-by-4 squared positions on the abdomen, providing sixteen signal records of the electro-physiological phenomenon measurable in slightly different positions of the abdominal wall. For future reference, the electrodes are numbered as illustrated in Figure 3.2a and their position in the patient's abdomen is illustrated in Figure 3.2b.

Records were captured with a 200Hz sampling frequency and a duration averaging 61 and 36 minutes for pregnancy and labour recordings, respectively. The database contains 122 recordings of 45 pregnant women , 10 of which took place during labor. Recordings durations vary from 19 to 86 minutes and gestational ages vary from 29 weeks and 5 days to 41 weeks and 5 days. Each case contains some information of the patient such as the patient's age, gestational and labor ages, previous pregnancies information, oxitocin administration and labor procedure informations.

A greater detail on the description of electrodes placement configuration, test protocol procedures, some of the decisions rational and the files they provide may be found in the database website [3].

This database seems a good choice for EHG studies as it has a standardized protocol, which is a requirement when analyzing signal characteristics that should not be influenced



(a) Electrodes 4x4 arrangement [3]

(b) Position of the electrodes arrangement in the patient's abdomen [3]

Figure 3.2: EHG Electrodes Placement [3]

by different recording methods and procedures. Furthermore, it has a fairly defensive protocol that may accommodate future discoveries. An example is the use of a 4 by 4 electrode grid, allowing study and usage of signal propagation characteristics. Also, it uses a recording sampling frequency of 200Hz to study a signal for which related studies consider a frequency band of interest going up to 2Hz, accommodating findings on relevant information for higher frequency components.

3.1.2 Uterine Physiological Events

Statistical approaches using EHG records pursue recognizing patterns in the evolution of contractions, among other physiological events. In this regard, it is hypothesized a use in the study of individual and the overall behavior of these events to assess the state of pregnancy.

A few uterine events of interest can be identified from the EHG signals, which include fetal movements and the contraction types: low amplitude high frequency (LAHF), Braxton-Hicks and long duration low frequency (LDBF) [49].

LAHF contractions or Alvarez waves, first described by Alvarez and Caldeyro [8], have intensity lower than 5mmHg, occur 1 to 3 times per minute, starting at the 9th week and occurring until the end of pregnancy. In their study these contractions are described as uterine fibrillation for their unsynchronized, localized characteristics, occurring randomly in different parts of the uterus [1, 10]. A study on the significance of Alvarez waves by Newman et al. [1] hypothesizes the possibility of Alvarez waves being more frequent in patients who subsequently developed preterm labor, although the statistical value made the theory inconclusive. Other studies state the existence of no correlation between the occurrence of LAHF contractions and parity nor gestational age and the difficulty of taking any conclusion given that these account for 70 to 80% of the total contractions recorded in normal pregnancies [10]. While their occurrence is not proven to be an indicator of labor

proximity, LAHF activity seems related to high excitability and poor coordination of the uterine muscle and should be further studied [10].

Braxton Hicks contractions are described as painless, non-rhythmical contractions of higher intensity [8]. Their occurrence increases during pregnancy being very sporadic in the first 8 months and only acquiring a regular rhythm in the last 2 weeks before labor with 3 to 9 occurrences each 10 minutes [8].

LDBF contractions, have a duration of 2 seconds, up to several minutes in a band of 0 to 1Hz. These are associated with uterine hypertonus, meaning an excessive tension of the uterus without retrieving to a complete relaxation of the muscle, which causes fetal distress and may jeopardize fetal well-being [4, 19].

These events electrical activity is usually studied considering a band of interest around 0.2 to 1Hz (further analyzed in 3.2.2) and an amplitude ranging $100\mu V$ to $1.8mV$ [49].

3.2 Signal Preprocessing

Raw data collected from physiological events requires very specific domain-dependent analysis and preparation to allow a solid interpretation. Rather than uterine activity alone, the data acquisition mechanism record an electric potential difference, which contains information from many other sources. The procedures to analyze and create new representations of the signals can be qualified as signal processing.

Signal editing represents a pre-analysing procedure to detect and remove negligible or degraded signals segments, if not whole cases. It is, most often, made visually by experts and may be a crucial step to discard data that will corrupt any further analysis.

Signal preprocessing covers all procedures applied to improve a signals readability. These, should, mainly, pick the relevant information and find logical, intuitive ways to represent it. EHG related studies include procedures such as resampling, filtering, offsetting, de-noising, signal transformation, burst detection, etc.

The procedures applied require a knowledge of the signals nature and of the domain dealt with. As physiological activity, the recorded information is bound to have nonlinear characteristics [19, 50]. In addition, EHG records contain information from other sources such as maternal electrocardiogram, abdominal muscle electrical activity, respiratory movements, electrode-skin contact potential fluctuation, movement artifacts and baseline components [5, 35, 38], which should be considering on signal processing.

In the context of a classification, finding the processing techniques that build up the signal's information relevance is a crucial step, as it becomes part of the classification model itself. After a training set has defined the classification process all new signals that should be analyzed have to go through the very same processing to become contextualized inputs of the classifier.

Inspecting the raw signal represented in Figure 3.3, it becomes clear that it holds little to no meaning. At this stage, the signal is tilted due to electrodes movement and adaptation to the skin, and the polarization of the skin-electrode interface, which acts as

a battery due to sweat, and contains undesirable information from different sources (will be discussed in 3.2.2).

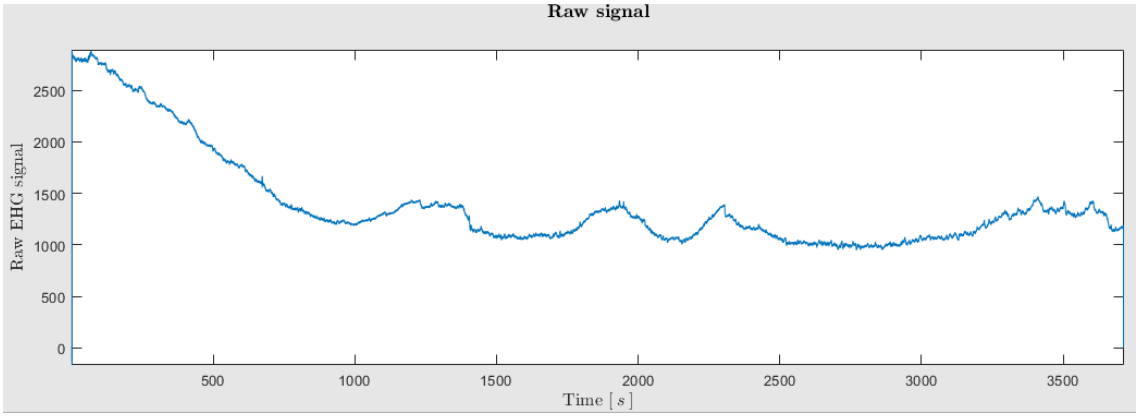


Figure 3.3: EHG Raw Signal. The amplitude has no meaning as it is relative due to machine gain.

Given the weak amplitude of uterine contractile events, ranging from $100\mu V$ to $1.8mV$ [49], the presence of multiple unbalanced sources and the different impedances in the skin, the recording mechanism requires an amplifier with a high gain to increase noise performance. As such, a first preprocessing procedure applied is dividing the machine gain to the raw signal to obtain a real amplitude scale.

It was observed high disturbances in the beginning and final parts of some test records. The disturbances seem to be caused by either starting the recording while still positioning or moving electrodes. To avoid these, 20 seconds were removed from the beggining and end of all records.

The remaining applied preprocessing techniques will be described in the following subsections.

3.2.1 Decimation

Given the contractile events have low-frequency characteristics, depending on the recorded sampling frequency an usually useful step to decrease computation time and signal length is down-sampling [38].

The band of interest of EHG signals is discussed in the next subsection 3.2.2, yet given that studies, usually, only consider frequencies as high 1-5Hz and recorded EHG sampling frequency takes values such as 100Hz [45], 200Hz [15, 18, 23, 31, 47] 500Hz [5, 13, 48], up to 1024Hz [38], decimation is usually possible and with great benefits. The reason sampling frequencies take such high values is only attributed to the fact that EHG has not been studied enough to determine that no high frequency components may be relevant. The price and effort required to create an EHG database motivates a defensive approach of performing an exorbitant measurement.

Related studies apply decimation to obtain sampling frequencies of 3.125Hz [51], 6.25Hz [18], 8Hz [41], 16Hz [38], 20Hz [44], etc. The most widely used criteria to define an appropriate sampling frequency is through Nyquist–Shannon sampling theorem, which states that, given a frequency f_s, perfect reconstruction is guaranteed possible for a bandwidth $B < \frac{f_s}{2}$. As such, decimation may be dimensioned through this criteria in a way that keeps intact the band of interest in the study of EHG records.

In this work, it will be considered that the uterine electrical activity band of interest is defined as 0.1Hz to 1Hz (which will further be discussed in subsection 3.2.2). Considering this band and according to the Nyquist–Shannon sampling theorem, the minimum sampling frequency that can be considered without loss of these components is 2Hz. However, as a defensive choice the sampling frequency was defined as 4Hz in order to minimize aliasing effect.

The EHG records found in the PhysioNet Database [3] are sampled at a frequency of 200Hz, as such a decimation filter was applied to the signal to re-sample it with a decimation factor of 50. The effect of the decimation is illustrated in Figure 3.4.

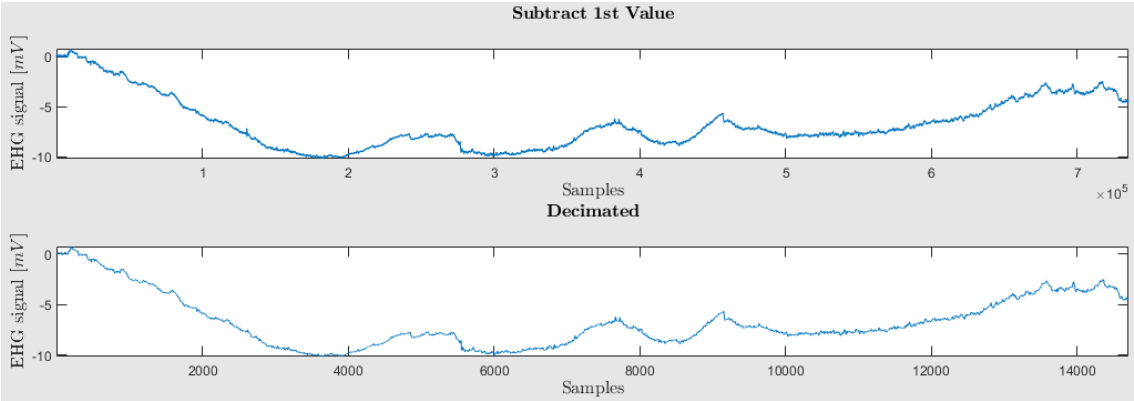


Figure 3.4: Effect of decimation. The plots represent a signal before and after applying the decimation filter, respectively.

As illustrated, the decimation performs a sample-wise rescaling of the signal. The signal size was reduced from roughly 8×10^5 samples to roughly 16000, which represents a substantial reduction of computation requirements for further use of the data.

3.2.2 Filtering Process

Digital signal filtering is a process that can be used to separate or restore signals. Signal separation is performed on signals that contain noise and interference from other signals. As an example, it should be used o the EHG records to extract uterine activity related events, while canceling other physiological phenomena such as the ones mentioned in the beginning of this section 3.2. Signal restoration is applied on cases where the signal has been distorted to attempt recovering the real source of the signal.

The signal separation process of extracting the desired frequency components of a signal and suppressing partial or completely the undesired ones may be performed by

either windowing the signal in the frequency domain or through discrete-time system operations in the time domain [52]. In short, the later may refer to one of two filter classes: Finite Impulse Response (FIR), carried out by convolution of the input signal with the filter's impulse response, and Infinite Impulse Response (IIR) filters, which are an extended version, with recursive characteristics [53]. As the name implies, the classes differ in their response to an impulse. Since FIR only uses present and past input information to determine respective output, it has a finite response to an impulse. Conversely, IIR, aside from input information is fed previous output information in it's estimates, which creates an infinite response [53].

When filtering EHG records, it is intended the separation of the uterine related events from the other physiological events recorded. To achieve this, there is some range of frequency bands considered. Amongst studies, there is, in the most part, an agreement in considering a range from 0.2Hz up to 1Hz [33, 34, 41, 48, 54]. Although, commonly used minimum frequencies are also 0.1Hz [5, 18, 39, 54] and 0.34Hz [14, 22, 39, 40, 42, 46, 51], which avoids respiration interference [14, 46]. Maximum frequencies considered may go up to 4Hz [5, 30, 31] and 5Hz [39].

The bands of choice is based on the fact that uterine contractile activity energy is mostly concentrated in a range up to 1Hz, along with the necessity to exclude the interfering artifacts. The main electrical signals considered interference and noise include:

- Maternal electrocardiogram (ECG): Electrical activity of the patients heart creates interference with a frequency band of main components ranging from 1.38Hz to 1.5Hz [4];
- Maternal respiration: Has a frequency band that ranges from 0.2 to 0.34Hz, being the most interfering component [4].
- EMG noise: Due to abdominal muscle activity recording. It's dominant component revolves around 30Hz [4];
- Motion: Has a frequency range between 1Hz and 10Hz and causes distortion due to relative motion of the electrodes and the decrease in muscle related to it's activation [55];
- Inherent noise in electrical equipments: Persistent, stable noise dependent of the high quality of the device with frequency components ranging 0Hz to several thousands Hertz [55];
- Ambient noise: The main source is the electromagnetic radiation associated with the power line, with frequency components of 50Hz or 60Hz [55];
- EMG inherent instability: Due to the unstable nature of motor units firing rate, with frequency components ranging 0Hz to 20Hz [55];

In this work it will be considered a band of interest ranging from 0.1Hz to 1Hz. This decision is based upon the spectral characteristics of the interfering sources previously listed and the fact that most of the uterine electrical activity energy is in that band [19, pp. 27]. In this regard a wavelet packet (WP) filter was applied to the signal, which was earlier developed in the UEX projectcontext and is described in [4]. This filter was chosen for it's stability, computational efficiency and simplicity.

Figure 3.5 shows the effect that the filter has in the EHG signal.

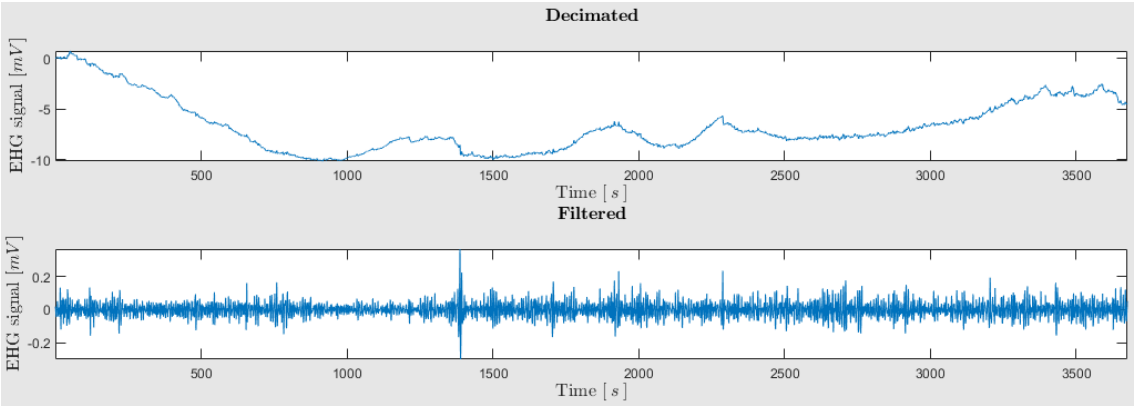


Figure 3.5: Effect of applying a pass-band filter with cutoff frequencies in 0.1 and 1Hz. The plots represent the signal before and after applying the filter, respectively.

At this point, those components outside the frequency band of interest are greatly attenuated it producing a more interpretable and relevant signal to the study of uterine events. Thus, the obtained signal will be smoother with reduced higher frequency noise and abrupt changes. In the other hand, the lower frequency components that produced the huge offset wave across that tilts the whole signal were attenuated producing a very clear result in 3.5.

3.2.3 Bipolar Signals

A mostly adopted method to reduce noise is the use of bipolar signals obtained by subtracting signals recorded form pairs of electrodes. Use of bipolar signals allows canceling common mode interferences and has been demonstrated to reduce a large portion of the noise [5, 38].

Evidently, subtracting pairs of signals to reduce noise will also have an effect on the uterine events and it has also been shown to have low spatial resolution when localizing and differentiating multiple dipole sources. The obtained signals are affected differently, as conductivity is not homogeneous among the electrodes, which affects the study of signal propagation [5]. The article [5] describes Laplacian bioelectric potential as a possible solution to the previous problem. Another additional procedure to the use of bipolar signals may be to their normalization by dividing standard deviation to each signal [25].

The method is claimed to ensure equal significance of features in the process of classification [25].

In this work, bipolar signals will be used with the same configuration as that of Catarina Sousa's work [56, pp. 52-55], whose contribution to UEX project included a study of the referred configuration. In their work, the reader may find a more detailed explanation of the decision made and a graphical example that shows the significant reduction of the maternal ECG influence in a bipolar signal compared to those of the used monopolars pair.

Computation of bipolar signals will be performed by the difference of monopolar signals respecting the positions illustrated in Figure 3.6 in accordance to the same numbering and position as in Figure 3.2a.

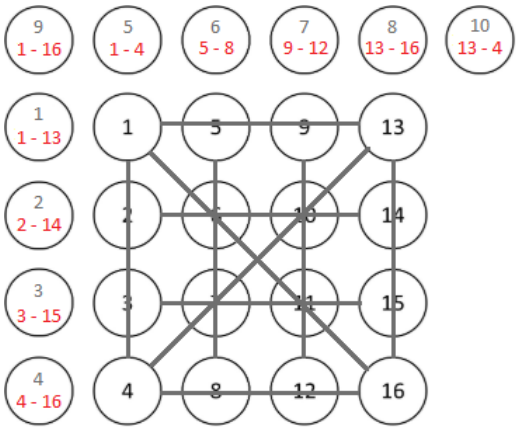


Figure 3.6: Relation of monopolar to bipolar signals. The top-most and left-most circles represent the calculated bipolar signals, which specify their respective numeration and the computed subtraction.

The Figure 3.6 shows how 10 bipolar signals are extracted from the subtraction of monopolar records corresponding to electrodes on opposite vertical, horizontal and diagonal positions. The distance between electrodes that define a bipolar signal was maximized to allow a good compromise between common noise reduction and avoiding canceling out relevant components in accordance to the results obtained by Catarina Sousa's work [56].

As such, by the configuration observed in Figure 3.6, bipolar signals are given by

$$
\begin{aligned}
EHGb_1 &= EHG_1 - EHG_9 \\
EHGb_2 &= EHG_2 - EHG_{14} \\
&\vdots \\
EHGb_5 &= EHG_1 - EHG_4 \\
&\vdots \\
EHGb_9 &= EHG_1 - EHG_{16} \\
EHGb_{10} &= EHG_{13} - EHG_4
\end{aligned}
$$

An example of the results obtained in this procedure represented in Figure 3.7.

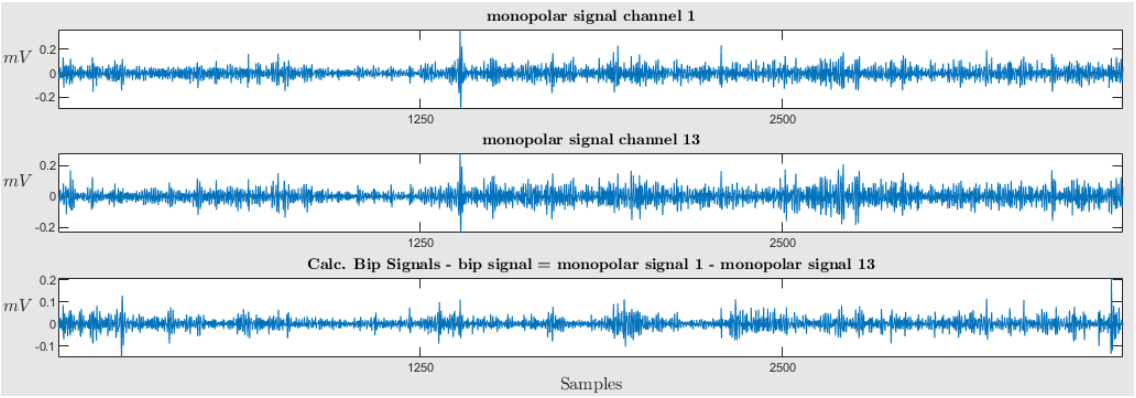


Figure 3.7: Effect of subtracting monopolar signals. The first two graphs represent records from different electrodes, while the third represents the bipolar signal resultant of the subtraction.

Figure 3.7 shows a pair of monopolar signals that show great similarities and the resulting bipolar signal. The later has a thinner base line and more visible events. Thus, it may be concluded that the procedure may have ominous influence over uterine events through their attenuation or manipulation, but also reduces common noise interferences and results in a cleaner signal with better distinguishable events.

3.2.4 Extraction of Uterine Events

Having a seemingly satisfactory manipulation of the test records with respect to noise reduction, ease of computation and enhancement of the components that characterize uterine events, the next step is extraction of the uterine events from the records. Optimally, this procedure should isolate uterine events consistent to those specified in 3.1.2, by scanning the signals for those characteristics that define each of those events.

Although there is some knowledge of how those events should be and what characterizes them, the analysis of real EHG records brings less defined boundaries, bringing some disagreement to the criteria used for extracting the contractile segments. The methods must define what constitutes a contraction along with where it begins and ends. Furthermore, the decision should not only define contractions with respect to baseline behavior, but also noise components, which can even lead to greatly represented outliers.

As such, it is first presented some studies that also performed contraction extraction, the methods they used and some of their considerations:

- Manual [15, 17, 18, 23, 25, 35].
- Algorithm based [30]:
    1. determination of EHG slow wave which represents the strength of contractile activity;

2. determination of basal tone which corresponds to the resting potential
3. application of detection level above basal tone to detect the candidate episodes of increased activity
4. validation as contractions when duration and amplitude related to the basal tone exceeds established minimal values;

- Manual selection according to the criteria [5]:
  - Significant rise in signal amplitude with respect to basal period;
  - Duration longer than 30s;
  - Signal morphology typical of electrophysiological changes. Signals with over-abrupt changes, saturation or coincidental in time with movements of the patient during the recording session are discarded.

- Algorithm based [39]:
  1. Calculate RMS and the unnormalized first statistical moment of the frequency spectrum of 30s moving windows displaced every 0.25s;
  2. identify segments with amplitude significantly different to that of the baseline;
  3. Identify Baseline activity with a 4 minute moving window displaced every 0.25s, where the baseline is assumed to be the average of the 10% lowest amplitude values;
  4. Identify Segments of interest that respect the conditions
     - Having over two times the mean baseline activity;
     - Having over 25% the signal amplitude of each window for more than 30s.
  5. classified by two experts as being or not an artifact.

- Manual selection where bursts were identified by [40]:
  - succession of increased-amplitude voltage spikes whose mean amplitude remained at over 2 times the mean baseline activity;
  - duration was over 10 seconds;
  - correspond, at least temporally, to uterine pressure events as measured by TOCO.

- Manual segmentation [33]:
  - Term labor group:
    - term labor group: contraction period consisted of 10 seconds before and after the peak of burst, summing 20 seconds;

    * select non-contractions from the middle of adjacent contractions also considering 20 second duration segments.

  – Non-labor group:

    * Does not consider a contraction period and every 20 seconds was selected manually as one non-contraction period.

This work will be using contraction extraction algorithms developed for the UEX project [4] by other contributors. The algorithms define candidate contractions through study of the the signal's energy estimation peaks and determines whether it is a contraction and it's boundaries through a peak prominence criteria, assessing it's amplitude and relative location to other peaks. Energy limits are also defined as a way to rule out outliers. The algorithm was implemented to consider five different energy estimators, including:

- Wavelet Energy (Figure 3.8);
- Teager Energy Operator (Figure 3.9);
- RMS (Figure 3.10);
- RMS squared (Figure 3.11);
- Hilbert-Huang Energy Spectrum (Figure 3.12);

The resulting contraction segments using each of the energy estimators of the algorithm are represented in Figures 3.8, 3.9, 3.10, 3.11 and 3.12.

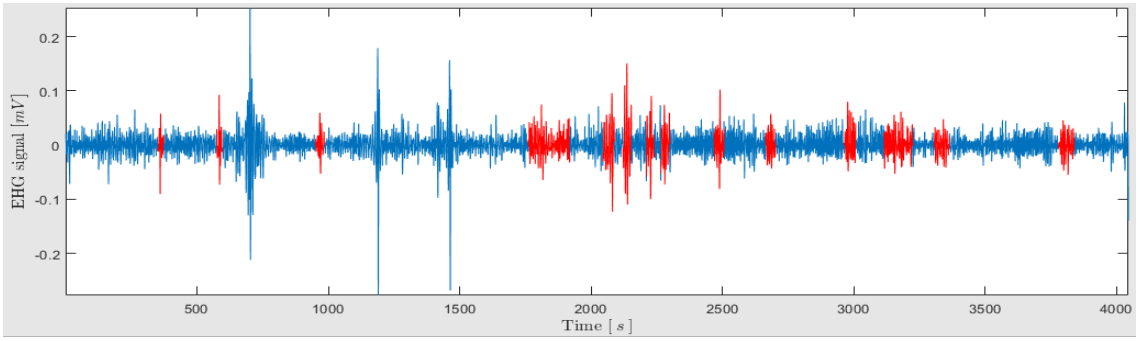


Figure 3.8: Detected contractions using Wavelet energy

It becomes clear from inspecting the Figures 3.8, 3.9, 3.10, 3.11, 3.12 that there is quite a big difference in the segmentation decisions of each method while considering the same signal. Results show that further work is required in defining the criteria for segmenting contractions or the robustness of each of the current ones. However, the inspection of which of these methods may provide the most rigorous contraction detection, as well as a more detailed explanation of how they are performing this task is out of this work's scope. The implementation of the classifier and the processing procedures to attain the input data will assume any of the methods may be used or that further methods may be

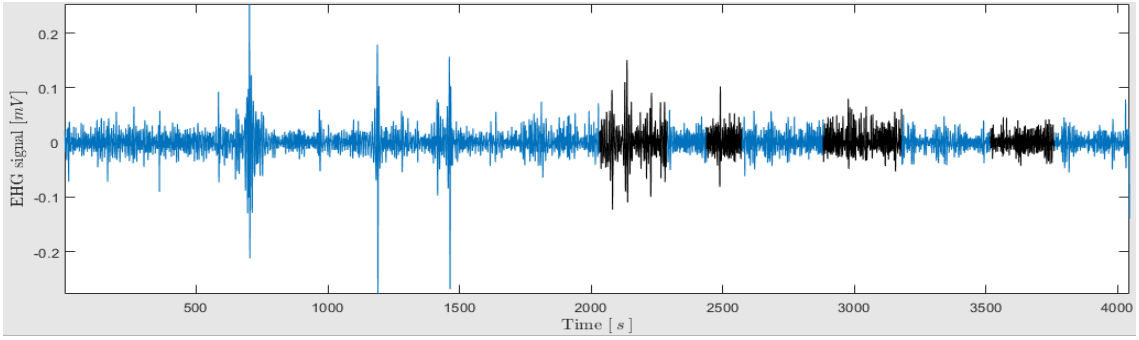


Figure 3.9: Detected contractions using Teager-Kaiser energy

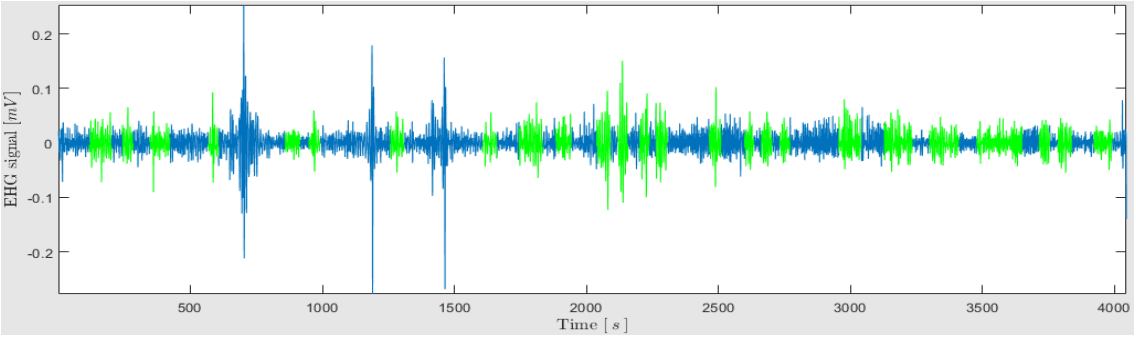


Figure 3.10: Detected contractions using RMS

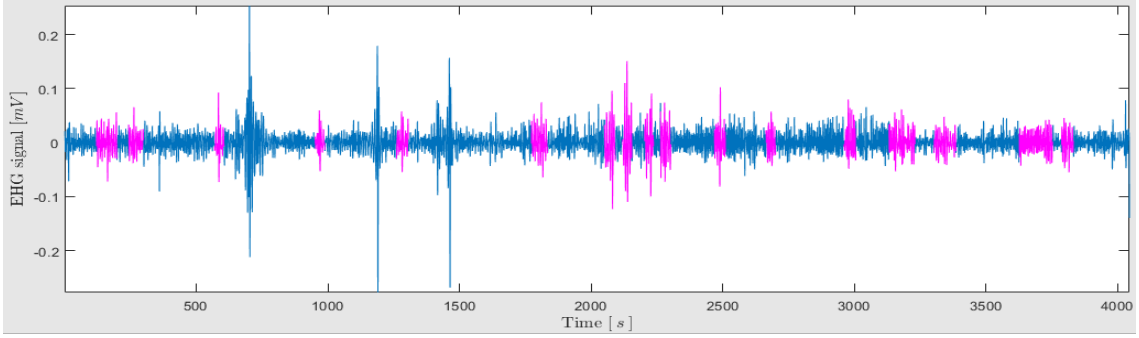


Figure 3.11: Detected contractions using RMS squared

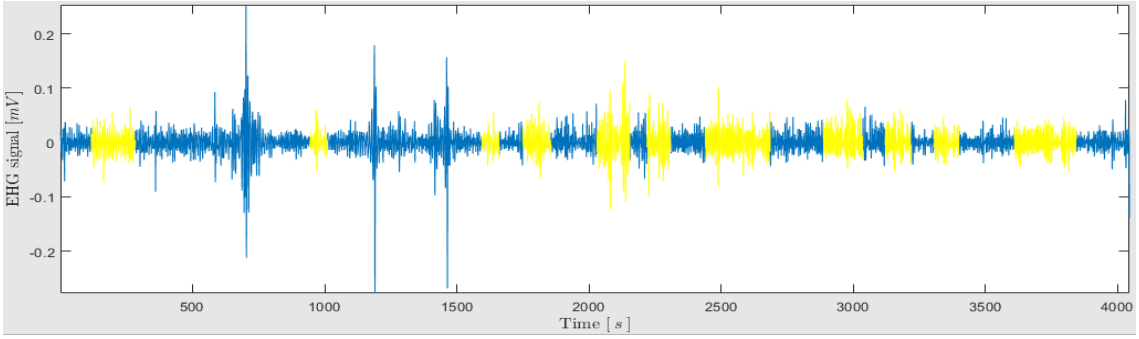


Figure 3.12: Detected contractions using Hilbert-Huang spectrum energy

added. In most of the upcoming tests wavelet energy method will be used as a standard, being the one that has shown the best results according to daily experience from other UEX project contributing members.

Another important conclusion from the figures is that, although the algorithm is based on peak detection it is visible that a few peaks were not considered contractions, which is a result of the outlier detection mechanism.

In these algorithms, fetal movements are extracted separately as they compose events of lower energy that would less probably be captured by the same standards as the other uterine events. That being the case fetal movements represent the only class that is already defined, prior to classification.

3.2.5 Signal Modeling

Signal Modeling refers to techniques that generate a prediction of the signal, in order to model it. There are a number of modeling methods, which may be useful by attempting to provide an equation that fits the signal [52].

Autoregressive model (AR) offers is a linear approach to signal modeling through the estimation of an output dependent on the current and previous inputs, along with feedback values. It, thus, makes the assumption that data points are closely related to the previous few data points [52]. According to a study comparing AR model and statistical classification methods on EHG signals [28], being EHG records the result of filtered elementary activities integration and given that filtering induces a correlation between successive samples, an AR model is made possible. Furthermore, the study argues that the use of an AR model allows the detection process to be applied to the prediction error rather than the original samples [28].

The AR model of a discrete process $x(n)$ of order q is defined as the output of a recursive filter whose input is a white noise process [28]:

$$x_n = \sum_{i=1}^{q} a_i x_{n-i} + \eta_n,$$

where η_n are the AR model's innovations [28].

Some of the AR models applied on on studies regarding EHG records are indicated bellow:

- Method based on minimization of the prediction error power $\sigma^2 = E(\eta_n)^2$, within an AR model of order 16. The AR model was then used to generate two types of simulated uterine contractions by using White Gaussian Noise (WGN) [28];

- Method based on Linear Prediction Filter Coefficients (lpc), which uses the autocorrelation method of AR modeling to find the filter coefficients. The AR model used is of order 16 to estimate AR coefficients of two types of EMG signal: those for which the pregnancy resulted in term delivery and for the ones that resulted in preterm

delivery. The mean average was then calculated for the lpc coefficients and of each group and two types of simulated contractions were generates by using WGN [2].

The brief analysis of signal modeling comes from the extensive use of modeling procedures in EHG studies [2, 28, 31, 49, 54], the fact that it is a very powerful tool in signal processing and for the fact that autoregressive models are used as optional methods for the estimation of contractions spectral representations in the UEX project [4].

3.2.6 Spectral Analysis Methodologies

An important step towards finding patterns is analyzing cycles in the signal. The most straightforward way to do so, is applying a transformation of the signal to the frequency domain, which provides the energy distributed among different cycles. Many methods can be applied to approach this problem, which will be discussed in the next subsections.

In general terms, there are four different ways to express the spectral representation of an input signal measured through a difference of potentials (Volts):

- Power Spectrum (PS): Representation of which frequencies contain the signal's power in the form of a power value distribution, expressed in V^2, as a function of frequency.

- Power Spectrum Density (PSD): Represents the power per cycle distribution as a function of frequency and can be defined as the Fourier Transform of the autocorrelation sequence r_x of the time series, expressed in $\frac{V^2}{Hz}$ [57, pp. 7]. Thus, it may be defined by

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-jk\omega} \tag{3.1}$$

  where r_x, for an ergodic process, is given by

$$r_x = \lim_{N\to\infty} \left\{ \frac{1}{2N+1} \sum_{n=-N}^{N} x(n+k)x^*(n) \right\} \tag{3.2}$$

  with x^*, as the complex conjugate of x. The PSD, in regard to long-duration signal segments, provides more accurate power delegation to the correct frequencies and reduces noise-induced fluctuations in power amplitudes at the cost of reducing frequency resolution, as fewer data points are available for each FFT calculation.

- Linear Spectral Density (LSD): The square root of the PSD, being expressed in $\frac{V}{Hz}$ defined as

$$LSD = \sqrt{PSD}$$

  The linear Spectra may be used for scaling purposes when it is necessary to provide a linear, direct relation with amplitude.

- Linear Spectrum (LS): The square root of the PS, expressed V. Similarly, provides a linear scale of power spectrum.

These different representations are all related to each other and a few methods will be covered to obtain each of them along this section.

3.2.6.1 Discrete Fourier Transform

Perhaps the most known and applied transform is the Discrete Fourier Transform (DFT), which deduces a representation of the signal in the frequency domain through a description of how the signal is constructed from the summation of sinusoids of different frequencies [31]. Considering a signal $x[n], n = 1, ..., N$ the DFT is defined as

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-j2\pi kn}{N}}$$

where n refers to time domain sampling, k to that of frequency domain. The inverse transformation can be obtained by the inverse DFT defined by

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W_N^{-kn}$$

with $W = e^{\frac{-j2\pi}{N}}$ and N as length of the signal. Fourier coefficients $X(k)$ are complex numbers, which can be represented in the Cartesian or polar forms

$$X(k) = X_{Real}(k) + jX_{Imaginary}(k) = |X(k)|e^{j\sigma}$$

Where X_{Real} and $X_{Imaginary}$ represent the real and imaginary parts in the Cartesian representation, respectively. The operation $|X(k)|$ refers to the polar vector's magnitude, while σ refers to it's angle.

Fourier transform (FT) represents a very simple approach to transform the signal. However, since it sums all frequency components over the whole observation interval, providing a single spectrum to the whole signal, it has poor time resolution [48, 52] and it's coherence and correlation rely on the measured signals stationarity [35], which constitutes a problem in biological signal analysis.

Another problem with the DFT is the computation requirements to calculate a sum of N multiplications with complex exponentials, which dramatically increases for large N. A solution to this problem is Fast-Fourier Transform (FFT) algorithms, which increase DFT performance by decomposing the DFT into DFTs of smaller length. Optimally FFT decomposes any N-point DFT into multiple 2-point DFT, however this is only possible for $N = 2^\nu$, with $\nu \in \mathbb{N}$. The optimal behavior if provided by the removal of redundant or unnecessary operations of the DFT.

3.2.6.2 Welch method

Welch method [58] estimates the power spectrum (PS) by dividing the time signal into successive blocks Assuming an input signal $x(n)$ to have finite energy, such that $n = 0, ..., N-1$. Welch method [58] estimates the power spectrum (PS) by dividing the

time signal into successive blocks of length L with an overlap that defines the segment starting points separation to be of D samples, resulting:

$$\begin{aligned} x_1(n) &= x(n), && n = 0, 1, ..., L-1 \\ x_2(n) &= x(n+D), && n = 0, 1, ..., L-1 \\ &\vdots \\ x_K(n) &= x(n+(K-1)D), && n = 0, 1, ..., L-1 \end{aligned}$$

where K refers to the number of segments that comprehend the entire input signal. This method, then, applies a window $W(n)$ to each segments, with $n = 0, 1, ..., L-1$, composing the sequence $A_i(n) = X_1(n)W(n), \ldots, X_K(n)W(n)$ with $i = 1, ..., K$. Applying the finite Fourier transform to the sequence it is obtained

$$A_i(k) = \frac{1}{L} \sum_{n=0}^{L-1} x_i(n) W(n) e^{\frac{-j2\pi kn}{N}}, \;\; i = 1, 2, ...K \tag{3.3}$$

The power spectrum is, then given by

$$I_i(f_n) = \frac{L}{U} \left| \sum_{n=0}^{L-1} x_i(n) W(n) e^{\frac{-j2\pi kn}{N}} \right|^2, \;\; i = 1, 2, ...K \tag{3.4}$$

with $f_n = \frac{n}{L}, \;\; n = 0, ..., \frac{L}{2}$ and

$$U = \frac{1}{L} \sum_{j=0}^{L-1} W^2(k)$$

Power spectral density is given by the segments periodograms (3.4) average, defined as

$$\hat{S}^W(f_n) = \frac{1}{K} \sum_{i=0}^{K-1} I_i(f_n)\,, \quad i = 0, \ldots, K-1 \tag{3.5}$$

The choice of the window function is a compromise of side-lobe level in the PSD estimate and frequency resolution. When the window is rectangular, overlap is meant to avoid loosing end-of-segment contribution. For generalized Hamming windows a common choice may be

$$\begin{cases} D = \frac{L}{2}\,, & L \quad even \\ D = \frac{L-1}{2}\,, & L \quad odd \end{cases}$$

whereas for Blackman window it is usually $D \approx \frac{L}{3}$.

3.2.6.3 Time-Frequency Analysis

Spectral analysis over the frequency domain has the drawback of losing time resolution, as it doesn't account for time-varying features [59]. This becomes a critical problem when considering biological signals for their non-stationary behavior.

Short-time Fourier transform (STFT) is a FT-based method, which provides an approach to the problem of time resolution. This method applies a window function centered in a time t to the signal and calculates the FT of the result. It then repeats the process while sliding the window through the signal. The window function should attenuate the signal near the ends of the window in order to avoid introducing discontinuity in the window, which is interpreted as an abrupt change, and to reduce spectral leakage caused by frequency components with cycle times higher than the window considered.

STFT is a powerful method on time-frequency analysis, however the fix sized sliding windows don't always account for the difference in duration of high and low frequency patterns, which makes it hard or impossible to defined a compromise between time and frequency resolution [59].

As a solution to the later problem, wavelet transform applies different sized windows in time to extract different frequency components [59]. This method makes use of a dynamic window formed of a function, denominated mother wavelet, which either is a sinusoid tapered to be compressed and stretched in time or uses changing scales.

Time frequency distributions (TFDs) are introduced as a means to study signals with varying signal frequency over time and neither time nor frequency representations accurately describe the signal [35]. Most used TFD estimation methods include Wigner-Ville distribution (WVD), spectrogram using a Short Time Fourier Transform (STFT), wavelet transform (WT) and Hilbert Huang transform (HHT) [35].

3.2.7 Spectral Analysis of Uterine Events

Spectral analysis was defined in subsection 3.2.6 as a way to analyze the cycles present in the signal. Intensive analysis of the spectrum and spectral density of EHG signals has been perform over previous works related to UEX [4] project in FCT. The same functions have been applied here to compute PS, PSD, LS and LSD, of the attained contractions and fetal movements, from different methods, namely:

- FFT;
- FFT Windowed (referred as fft-wind in Figure 3.13);
- Welch;
- Covariance (referred as cov in Figure 3.13);
- Burg;
- Modified Covariance (referred as mcov in Figure 3.13);
- Yulear;
- AR;

An example of the PSD calculated for a contraction, by the different methods is illustrated in Figure 3.13.

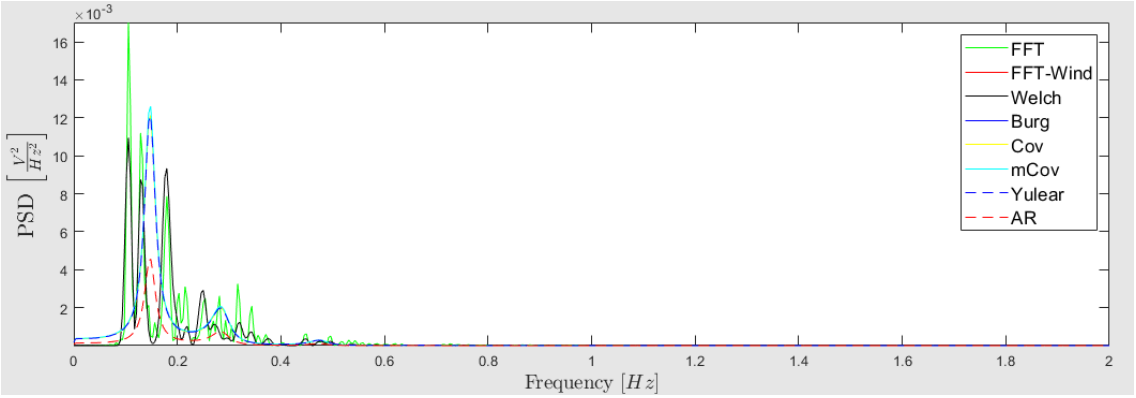


Figure 3.13: PSD representation of a contraction by different methods

Although, no distinction will be made in the implementation of this work, further tests will be using PSD as the spectral representation method. PSD was chosen rather than PS for it's superior capacity to represent periodicities derived from assigning units of power to each cycle as a function of frequency. LS and LSD were not considered, as no reason to benefit from a linear proportion to signal's amplitude was found in the course of this work.

With regards to the method used for PSD estimation, in further tests it will be used Welch method. The later was chosen for offering a good compromise between representing the different frequency components and offering a stable solution. In other words, and inspecting Figure 3.13, Welch method it is an intermediate solution between the fft that shows a very fluctuating result with many peaks and the opposite results provided by all other methods that provide a two peaks representation of the signal.

As described in 3.2.6.2, Welch method requires a window that slides across the signal to perform the spectral estimation of successive blocks. The most commonly used windows ω applied to general filtering problems are represented in the below list:

- Rectangular

$$\omega(n) = \begin{cases} 1\,, & 0 \leq n \leq N \\ 0\,, & otherwise \leq n \leq N \end{cases}$$

  with window length N.

- Blackman window

$$\omega(n) = \begin{cases} \frac{2n}{N}\,, & 0 \leq n \leq \frac{N}{2} \\ 2 - \frac{2n}{N}\,, & \frac{N}{2} \leq n \leq N \end{cases}$$

  with window length $L = N + 1$.

- Hanning window

$$\omega(n) = 0.5\left(1 - \cos\left(2\pi\frac{n}{N}\right)\right)\,, \quad 0 \leq n \leq M - 1$$

with window length of $N-1$

- Hamming window

$$\omega(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right), \ 0 \leq n \leq M-1$$

with window length of $N-1$

- Blackman window

$$\omega(n) = 0.42 - 0.5\cos\frac{2\pi n}{N-1} + 0.08\cos\left(\frac{4\pi n}{N-1}\right), \ 0 \leq n \leq M-1,$$

where N is the window length and $M = \begin{cases} \frac{N}{2}, Neven \\ \frac{N+1}{2}, Nodd \end{cases}$.

These windows have known behaviors, stated in table 3.1, which should be weighted as a number of trade-offs for application-dependent approaches.

Window specifications are usually seen through the frequency response in decibels (dB), given by

$$h_{dB}][f] = 20\log h_{Amplitude}[f]$$

The decibel unit describes a logarithmic ratio, providing the amplification or attenuation effect in a logarithmic scale is defined by the formula

$$ratio[dB] = 10\log\frac{signalpower}{referencepower} = 20\log\frac{signalamplitued}{referenceamplitude}$$

which, thus, returns a ratio value and no longer an amplitude or power. Given the above equation a calculated ratio of 20dB would be interpreted as an amplification of factor 10 to the signal, while -20dB would have a similar scale of attenuation.

The frequency response in decibels of Hanning, Hamming and Blackman windows is exemplified in figure 3.15

The most relevant values, while evaluating the earlier mentioned windows are given by the window's frequency response characteristics, which are provided, with $L = M+1$, by the following table [60]:

A procedure of designing a windowing filter may respect the following sequence:

1. Define filter specifications;
2. Select the window based on stopband attenuation;
3. Determine window length based on transition width;
4. Design ideal filter based on cutoff frequency;
5. Introduce causality through a delay $\alpha = \frac{M}{2}$;
6. Multiply window function.

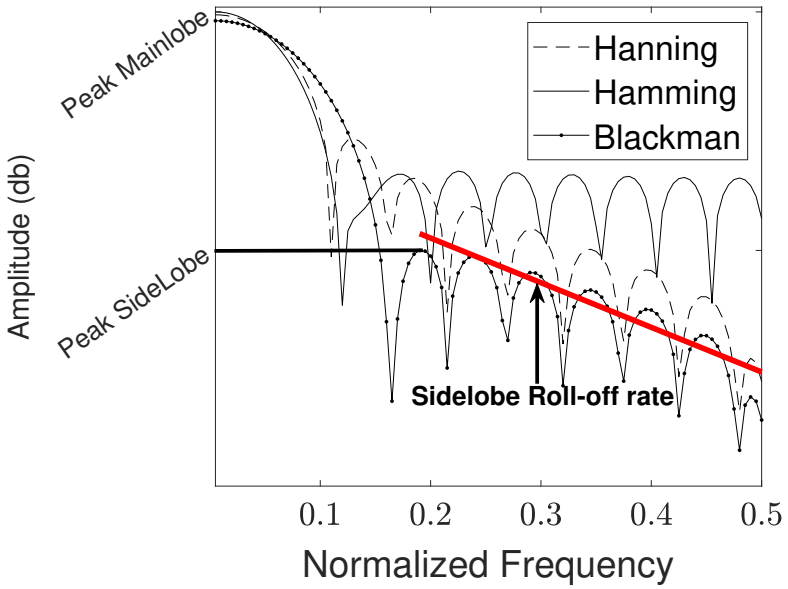


Figure 3.14: Frequency Response of Hanning, Hamming and Blackman Windows in decibels

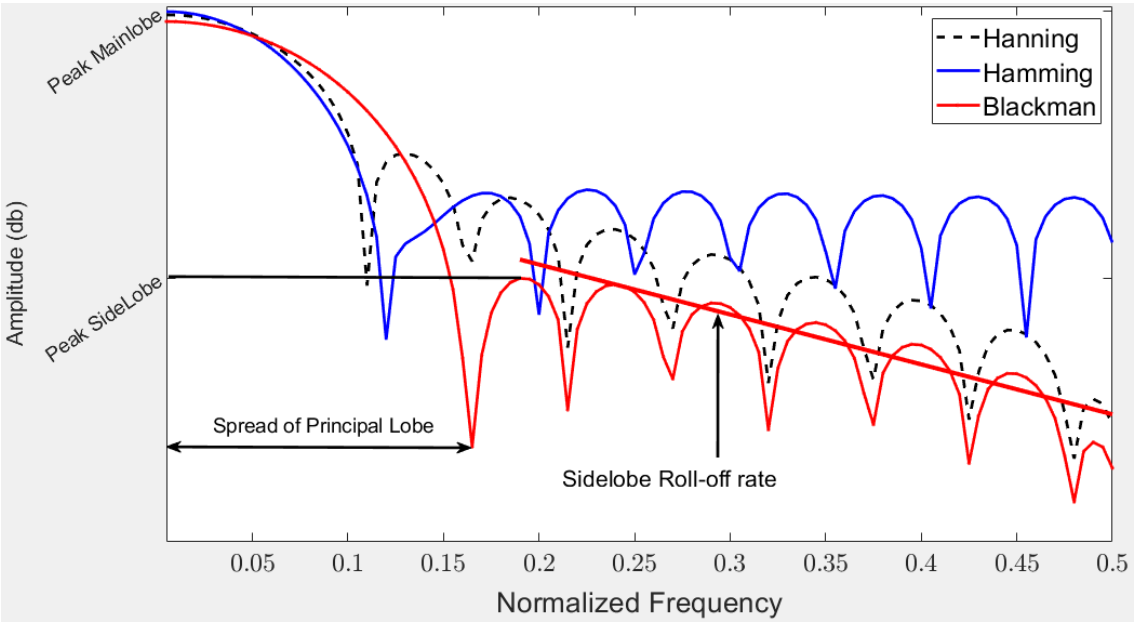


Figure 3.15: Frequency Response of Hanning, Hamming and Blackman Windows in decibels

Window	**Peak Sidelobe Amplitude (dB)**	**Approximate Width of Mainlobe**	**Peak Approximation Error (dB)**
Rectangular	-13	$\frac{4\pi}{M+1}$	-21
Bartlett	-25	$\frac{8\pi}{M+1}$	-26
Hanning	-31	$\frac{8\pi}{M+1}$	-44
Hamming	-41	$\frac{8\pi}{M+1}$	-53
Blackman	-57	$\frac{12\pi}{M+1}$	-74

Table 3.1: Characteristics of commonly used windows [60]

Welch method applied for further tests will be using a Hanning window as it provides a good compromise between frequency and amplitude resolution. As opposed to Hamming, which prioritizes frequency resolution, and Blackman, which has a very good roll off rate, avoiding power leakage, and has a good amplitude frequency, but ends up loosing frequency resolution.

Chapter 4

Dimensionality Reduction

Dimensionality reduction in pattern recognition concerns techniques designed to reduce feature complexity, computational requirements, remove irrelevant and redundant features, and reduces training required, which is highly dependent on feature dimensionality for most algorithms [61, pp. 112]. Although dimensionality reduction is a preprocessing technique, it is, here, considered in a separate chapter for the importance and detail it requires. There is no trivial answer to what best characterizes EHG signals, much less in the context of contraction unsupervised classification, which has, by definition, a foggy objective ahead. This makes dimensionality reduction solutions the most important step to attain relevant classes.

Approaches to dimensionality reduction problem can be divided into two categories, namely, feature extraction and feature selection.

4.1 Feature Extraction

4.1.1 Introduction

Features are values which, optimally, accentuate relevant characteristics of a signal. Hence, feature extraction is the procedure of deriving characteristics from the signals to create a better representation of their similarities and dissimilarities.

A book by Palaniappan on biological signals [52], describes features as computed values that should be representative of the signal and reproducible at different times. It describes as important criteria:

- lower dimension than signal
- high inter-class variance and low intra-class variance
- provide a robust and enhanced representation of the signal

Bassam Moslem et al. [25] bases feature selection on providing a meaningful electrophysiological interpretation of the EHG records, describing this procedure as domain dependent.

This section aims at reviewing features used by other studies, in regard to EMG records, with special focus those used on EHG. The features will be presented categorized by their linearity and domain characteristics. An attempt to identify each feature's meaning and context will be made in a way that allows an evaluation of whether it might be interesting in the specific context of this work.

Regarding the context of EHG recordings study, one important detail is whether contraction segments were used or the whole record. This choice will have an influence on features selection, as the relevant characteristics to profile a single contraction may differ from those that attempt to make sense of a whole record.

A concerning problem with directly studying the contraction segments is the non-linear characteristics of the uterine physiological events. These may imply better results from non-linear features over linear ones, as a way to represent this complex, non-linear system. The article [24], states non-linear techniques are less accurate when considering shorter amount of data, which becomes a problem given the time contractions take.

The later concern is further substantiated by article [30], which makes a comparison of the results obtained in the study of Fele-Žorž et al. [24] to detect preterm deliveries using the whole signals with theirs, while using the same features applied on contraction segments. The comparison shows that the non-linear features sample entropy and correlation dimension, were unable to distinguish data of term and preterm deliveries, achieving worse results with these features.

Nevertheless, the use of contraction segments allows other considerations such as standard time-domain descriptions of contractions [30] and one can hypothesize that it provides more specific information over relevant events.

4.1.2 Linear Time Domain Features

The most studied topic over EHG records is the prediction of preterm labor, for which the data is analyzed to identify patterns that differentiate preterm EHG records from normal pregnancies. Studies on preterm pregnancy may also try to find the same differentiation over labor records in comparison to those of pregnancy. In this regard, the selected features should outline a state of labor proximity.

The first group of features that will be examined are linear, time-domain features. These are listed in Table 4.1, with fields:

- Feature: Refers to the name of the feature;
- Context: Refers to the data used, more specifically, *'EHG signal'* refers to a study where the whole record was used, while the abbreviation *'EHG segm.'* refers to a study where the signal was segmented;
- Description: Attempts to describe the significance of the feature and provide studies comments on their results.

Feature	context	Observations
Amplitude	EHG segm.	– The article [30] found amplitude as incapable to distinguish term and preterm groups.
Area [30]	EHG segm.	– Area feature represents the sum of all points ($A = \sum_{i=1}^{n} x_i$), which implies a high sensitivity to artifacts.
Autocorrelation [24]	EHG signal	– In [24] it is described as a diagnostic tool for discriminating between periodic and stochastic behavior. It correlates the signal with a delayed version of itself, in order to identify patterns through periodic increased and decreased correlations.
Autocorrelation first-zero crossing	EHG signal [24] EHG segm.[30]	– time shift τ_1 of the first zero-crossing of the autocorrelation function after the autocorrelation peak occurs. It quantifies periodicy of action potential spikes within the burst [30]. It held good results in [24, 30].
Average duration of contraction [40]	EHG segm.	– Results show it reduced classification accuracy in study [40].
Average amplitude change [27]	EHG signal	– Sums the modulus of amplitude variations in the signal divided by the signal length. Thus, the feature represents how fast the signal is changing.
Average standard deviation of the burst duration	EHG signal	– Results show it was able to distinguish between term-labor, preterm non-labor, and term non-labor groups [40].
Contraction intensity [30]	EHG segm.	– average value of positive and negative peaks normalized into 1-min interval [30]. The article [30] found contraction intensity as incapable to distinguish term and preterm groups.
Difference absolute standard deviation [14]	EHG signal	– A eq deles n tem nada a ver com o desvio padrao pois n usa a media mas o ponto anterior.
Duration [30]	EHG segm.	– The article [30] found duration as incapable to distinguish term and preterm groups.
Kurtosis (K) [39]	EHG segm.	– Identify sudden large amplitude variations [39]. Results showed it was able to distinguish artifacts in study [39]
Log detector [14]	EHG signal	–
Maximum fractal length [14]	EHG signal	–
Mean absolute value ($MAVG$) [14]	-	
Mean of Total activity [40]	EHG segm.	– Results show it reduced classification accuracy in study [40].
Mean of burst duration [40]	EHG segm.	– Number of bursts in 30 minute period times the mean burst duration for that patient. Results show it reduced classification accuracy in study [40].
Normalized maximum derivative in relation to standard deviation of the base line [39]	EHG segm.	– Identify sudden large amplitude variations [39]. Results showed it was able to distinguish artifacts in study [39]
Normalized maximum derivative in relation to standard deviation of the EHG signal [39]	EHG segm.	– Identify sudden large amplitude variations [39]. Results showed it was able to distinguish artifacts in study [39]

Feature	context	Observations
Number of bursts per unit time [40]	EHG segm.	– Results show it reduced classification accuracy in study [40].
Ratio of segment RMS and baseline RMS [39]	EHG signal	– Identifies sudden large amplitude variations [39]. Was unable to explicitly distinguish artifacts in study [39]
Relative Amplitude [39]	EHG segm.	– Identify sudden large amplitude variations [39]. Results showed it was able to distinguish artifacts in study
Root mean square (RMS)	EHG signal [14, 24] EHG segm.[30, 33]	– A measure of signal power. Can also be a measure of deviation when signal has 0 mean. The article [30] found RMS as incapable to distinguish term and preterm groups. Study [33] found RMS significantly higher for contractions than non-contractions.
Standard deviation of burst duration [40]	EHG segm.	– Intends to measures variations in contractions duration. Showed the evolution on contraction effectiveness between labor stages, being able to distinguish labor [29]
Standard deviation [39, 40]	EHG segm.	– Exhibited bad or not explicit results in studies [39, 40]

Table 4.1: Linear Time-Domain Features

4.1.3 Spectral Features

In accordance to the spectral representation classes described in the beginning of this session 3.2.6 their use and acquisition methods by some of the considered studies are listed in Table 4.2.

Representation	Method
PSD	Welch Periodogram using window of type nfft, with size equal to half the length of the signal and 50% overlap for a total of 3 windows used. [15] FFT [2, 14, 24] FT [31, 62] Periodogram with a Hamming window [39]
PS	FFT [24] FFT periodogram (FFT-periodogram) [40] - Average of the periodograms calculated with FFT using Cosine-bell windows with 50% overlap. AR Model [31]. The study creates an AR model (see subsection 3.2.5) and estimates PS as defined by $$P_{AR}(f) = \frac{err_p^2 \Delta T}{\left\lvert 1 + \sum_{k=0}^{p} a_p(k) e^{-j2\pi k \Delta T} \right\rvert^2}$$ where ΔT is the sample period, err_p^2 is the sum of the squared forward prediction errors of AR model and order p was optimized by Akaike information criterion.

Table 4.2: Spectral Representations Considered in EHG Studies and Methods used

Similarly as in 4.1.2, the spectral features considered on some of the EHG records related studies are listed in Table 4.3.

Feature	context	Observations
Contraction power [30]	EHG segm.	– Measured as the sum of PSD components amplitudes multiplied by frequency resolution [30]. Results showed it was able to distinguish term and preterm groups according to article [30].
Deciles [2]	EHG segm.	– Decile method sorts data into ten equal parts: 10th, 20th, ..., 90th and 100th percentiles.
Energy for each frequency window [39]	EHG segm.	– Normalized according to total energy. Three frequency ranges are considered in PSD by using Hamming windows to attain energy distribution along the EHG record [39]. Results showed it was unable to distinguish artifacts in study
Maximum amplitude from analytic function [17]	EHG segm.	– The analytic function is obtained from HTT and HSA methods. With the high time-frequency resolution, HHT technology can better reveal the inner scales of signals [17]
Maximum power frequency [30]	EHG segm.	– Frequency component of higher. Results showed it was able to distinguish term and preterm groups according to article [30].
Mean frequency [15, 30]	EHG segm.	– [30] weighted arithmetic mean of PSD given as $$F_{mean} = \frac{\sum_{i=1}^{M} f_i PSD_i}{\sum_{i=1}^{M} PSD_i}$$ Results showed it was able to distinguish term and preterm groups according to article [30]. Study [33] results suggest it was relevant on identifying contraction and differentiating term and preterm groups.
Median frequency	EHG segm.[30], EHG signal[14, 24]	– Offers a way to capture some of the information of smaller peaks [24]. Study [24] identified slight drop in values (0.64 to 0.56Hz) as gestational age progresses. Results showed it was able to distinguish term and preterm groups according to article [30]. Study [33] results suggest it was relevant on identifying contraction and differentiating term and preterm groups.
Normalized Energy of PSD in windows $E_1 : 0.1 - 0.3Hz$; $E_2 : 0.3 - 1Hz$; $E_3 : 1 - 4Hz$ [39]	EHG segm.	– Provides the ration of energy distribution among the considered bands. Results showed it was unable to distinguish artifacts in study [39]
Peak frequency	EHG segm.[14, 15, 33, 40], EHG signal[24]	– Is linked to contraction strength [40]. Results show it was able to distinguish between term-labor, preterm non-labor, and term non-labor groups [40].
Power ratio [30]	EHG segm.	– Ratio of the PSD calculated in the band 0.8-3Hz and that of 0.25-0.54Hz [30]. The article [30] found power ratio as unable to distinguish between term and preterm groups.
Ratio of average power spectrum peak frequency and standard deviation of burst duration [40]	EHG segm.	– Was able to distinguish between term-labor, preterm non-labor, and term non-labor groups [40].
Standard Deviation of power spectrum frequency [40]	EHG segm.	– Results show it reduced classification accuracy in study [40].

Table 4.3: Non-Linear Time-Domain Features

4.1.4 Non-Linear Features

A study on the application of nonlinear time series methods in EHG segmented contractions and it's use in detection of labor [23] underlines the importance to justify the application of such methods by establishing the nonlinearity of the time series under investigation.

The study divides the methods to detect nonlinear characteristics in time series in [23]:

- methods inspired from chaos theory such as: maximal Lyapunov exponents, correlation dimension and transfer entropy. main disadvantage of these methods is that they depend on different parameters, like embedding dimension, which complicate the interpretations of the results
- predictability of time series methods: delay vector variance and approximate entropy. more robust than the first one but its interpretation is still not easy due to its sensitivity to the choice of the different parameters.
- statistical approach methods: time reversibility and higher dimensional autocorrelation functions.
- A combination of previously mentioned methods referring a new method called correntropy that computes a similarity index combining the signal time structure and the statistical distribution of the signal's amplitudes in a single function.

The previously mentioned study [23] makes special focus on the comparison of different nonlinear methods as to their applicability in EHG records. The study covers time reversibility, correntropy and approximate entropy methods as an approach to the problem. The statistical significance of each method is then tested by using surrogates. The study concludes time reversibility to be superior to the other methods. The results show uterine contractions during pregnancy to be reversible, whereas labor contractions to be temporally irreversible. Therefore, time reversibility obtained good results and should be a characteristic to consider in the study of EHG records.

Feature	**context**	**Observations**
Correlation dimension [24, 30]	EHG signal [24], EHG segm. [30]	– Estimates complexity of the time-series [24]. Held no discriminative results in [30], while promising results were found in [24]
Lyapunov exponent	EHG segm.[15, 30], EHG signal[24]	– Estimates amount of chaos in a system [24], studying system's stability and sensitivity to initial conditions [15, 30]. Held no significant results in study [24]
Sample entropy [14, 24, 29]	EHG segm.	– Measure of regularity of finite length time-series and estimates the extent to which the data did not arise from a random process. [24]. In [24] it is also mentioned that the results are highly susceptible to the parameter settings and create outliers. Suitable for complex and irregular time series signal to evaluate and distinguish the patterns [29]. Sensitive to noise, yet robust in obtaining significant evidence from the signal. Showed good results in [24]

Feature	context	Observations
Sample entropy [15, 30, 39]	EHG segm.	– Measures complexity and regularity of finite length time series, where a less predictable signal will present a higher sample entropy [15, 30, 39]. The article [15] determined the number of patterns to consider through the false nearest neighbor (FNN) method. Showed no discriminatory results, according to article [30]. Results showed it was unable to distinguish artifacts in study [39]
Time reversibility (Tr) [15, 23, 33, 39]	EHG segm.	– An indicator of non-linearity, as a time series is reversible if the probabilistic properties are unchanged with respect to time reversal [15]. Results showed it was able to distinguish artifacts in study [39]. Results of study [23] describe uterine contractions as reversible during pregnancy and temporarily irreversible during labor, along with the superiority of this feature in comparison with approximate entropy and correntropy. Study [33] identified a more variable time reversibility in contractions than non-contractions.
Variance entropy ($VarEn$) [15]	EHG segm.	– This method combines variance with sample entropy via inverse-variance wighting. Choosing the number of windows is a compromise between computation time and ability to detect variability [15].

Table 4.4: Non-Linear Time-Domain Features

4.1.5 General Feature Extraction Methods

Up to now, this section 4.1 has been focusing on feature extraction specific to EHG signals. However, while it has been discussed what features are usually considered, more general approaches to feature extraction will be revised here. These more general methods are exposed after reviewing EHG usual features, as they may be used as feature extraction methods applied on the contraction segments or their spectral representations, or even over an already defined set of features performing a feature selection task.

As aforementioned feature extraction aims at extracting a meaningful representation of our data, which not only decreases computation requirements for further processing of the data, but also to better influence the classifier. A few, quite similar solutions to this problem are principal component analysis (PCA), linear discriminant analysis (LDA) and intrinsic component analysis (ICA). As to illustrate their functioning it will be described PCA, in more detail.

Applied on dimensionality reduction context, PCA is a method that attempts to create a linear projection of lower dimensional subspace out of the initial set, using a variance maximization criteria. To achieve this, PCA performs an eigen analysis of the correlation matrix to obtain a geometric transformation to the set that minimizes it's statistical dependence.

To illustrate how PCA works, let us considering a set $x \in \Re^{M \times N}$, where M represents the number of observations and N the number of features and x_i are feature vectors with $i = 1, \ldots, M$. Before applying PCA the feature vectors are usually modified to have

zero-mean. However, when zero is a value of special significance, this modification may not be applied in order to preserve it's meaning [63, pp. 275]. Feature's mean vector $m_x \in \Re^{1\times N}$ is defined as

$$m_x(j) = \sum_{i=1}^{M} x_i(j), \quad j = 1, \ldots, N$$

The succeeding procedures assume zero-mean feature vectors are being used, given by the subtraction of the feature vector with it's respective mean value $m_x(j)$, however, if this not be the case $m_x(j)$ should be omitted from the proceeding equations.

On to applying PCA, it is first calculated the covariance matrix R_{xx} [61], defined as

$$R_{xx}(j,k) = \frac{1}{N-1}\sum_{i=1}^{M}(x_i(j) - m_x(j))(x_i(k) - m_x(k)), \{j,k\} = 1, \ldots, N \tag{4.1}$$

with $R_{xx} \in \Re^{N,N}$ resultant of the pairwise covariance between each column combination

$$R_{xx} = \begin{Bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \ldots & \sigma_{1N}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \ldots & \sigma_{2N}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1}^2 & \sigma_{N2}^2 & \ldots & \sigma_{NN}^2 \end{Bmatrix}$$

where the diagonal values represent column variances and the other entries $\sigma_{jk},\ j \neq k$ refers to the covariance of the column j and k. To obtain the linear transformation that fits the data, PCA performs an eigen analysis of the covariance matrix. As such, the eigenvalues and eigenvectors are calculated to satisfy the equation

$$R_{xx} = U\Lambda U^{-1}$$

where Λ is the diagonal eigenvalue matrix and U is the eigenvector matrix of R_{xx}. The eigenvectors can now produce a the linear PCA transform to our data with increasing relevance according to how high the correspondent eigenvalue is. As such, the eigenvector that corresponds to the largest eigenvalue is denominated the principal component. Having said this, the eigenvector matrix can now be sorted according to their respective eigenvalues and reduced to a number of columns that correspond to the highest ones. Later choice can be based on a plot of the sorted eigenvalues 4.1, where the iteration that marks the evolution from an abrupt descend to a flat line is identified as relevance limit (elbow technique). More specific methods include Akaikes's information criteria and the minimum description length criteria [64, pp. 182].

From this point, our initial feature set x, as well as any new values that respect the same context, can now be represented according to the linear transformation obtained

$$x' = U^T(x - m_x)$$

where x' refers to the new projection of x, m_x is maintained as how it was calculated while performing PCA and U^T is the transpose of $U \in \Re^{N\times L}$, being L the number of relevant

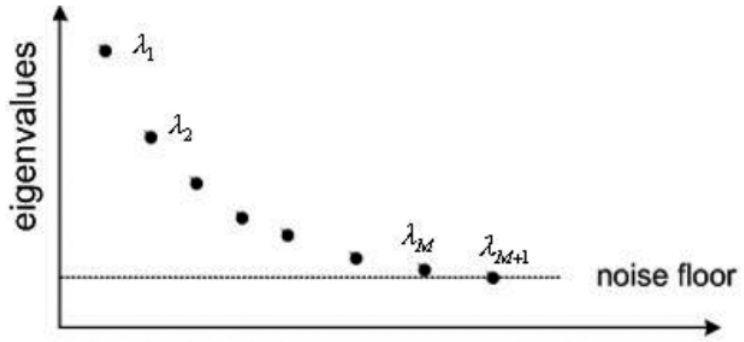


Figure 4.1: Determination of number of sources by eigenvalue ordering [64, pp. 182]

eigenvectors. Likewise, the inverse transformation is given by

$$x = Ux' + mx$$

This method is very powerful and has been used in the context of EMG, EHG and electroencephalography (EEG) by studies, such as those referenced in [18, 42, 59, 65].

Some of the feature extraction methods considered in EHG related studies include:

- Statistical significance [14];
- Linear discriminant analysis (LDA) using independent search (LDAi) [14];
- LDA using forward search (LDAf) [14];
- LDA using backward search (LDAb) [14];
- Gramschmidt (GS) analysis [14];

4.2 Feature Selection

Feature selection covers methods that aim at finding the most relevant subset of features out of the ones initially considered. These should be able to evaluate features discriminatory abilities, preferably, considering not only how they perform singularly, but also as a whole. Within this considerations, feature selection should not only assess the criteria listed in 4.1.1 of what a features should be, but also identify redundancies and overall misrepresentations associated with a subset of features. These methods, mainly, fall into three general categories: filter, wrapper and embedded approaches.

Embedded approaches cover techniques embedded in the training process, being designed to specific learning machines. This becomes a problem when considering the limitations entailed by the aim of this work, hence these techniques will not be covered here.

Filter and wrapper approaches will be covered in the next subsections.

4.2.1 Filter Approaches

Filter approaches rule out features that have little chance to be useful in analysis of data as a preprocessing step, independently of the classification or learning process. This, characteristics make it the most simple, generalized and computationally inexpensive approach. The criteria is based on a performance evaluation of the data, where an algorithmic procedure is used to estimate the relevance index [63, ch. 3 pp. 89-90 Wlodzislaw Duch]. Most filter methods, however, require either knowing the data classes, which would be the case for a supervised classification or à priori class probabilities knowledge.

An example of a class-independent filter approach is the Fast correlated based filter (FCBF). This method is based on symmetrical uncertainty, which is defined as the ratio between information gain IG and the entropy H of feature pairs [66]. Consider a dataset containing M features calculated for N observations, such that $x \in \Re^{M \times N}$ and x_i are feature vectors with $i = 1, \ldots, M$. The symmetrical uncertainty of the pair x_i, x_j is, then defined as

$$SU(x_i, x_j) = 2\frac{IG(x_i, x_j)}{H(x_i) + x_j}, i \neq j$$

with information gain and entropy defined as

$$IG(x_i, x_j) = H(x_i) + H(x_j) - H(x_i, x_j), i \neq j \tag{4.2}$$

where the entropy of a single feature is given by

$$H(x) = -\sum_{k=1}^{M} P(x(k)) log_b P(x(k)) \ , \ k = 1, \ldots, M$$

being b the logarithmic value, which commonly takes the value of 2,e (Euler's number) or 10, $P(x(k))$ is the marginal probability density function of a random phenomenon $x(k)$. This probability is implementation-dependent but as an example it can be considered

$$P(x(k)) = \frac{x^2(k)}{\sum_{k=1}^{M} x^2(k)} \ , \ k = 1, \ldots, M$$

In 4.2, $H(x_i, x_j)$ refers to the joint entropy of x_i and x_j, defined as

$$H(x_i, x_j) = -\sum_{k=1}^{M}\sum_{m=1}^{M} P(x_i, x_j) log_b [P(x_i, x_j)]$$

where $P(x_i, x_j)$ refers to the joint probability of both events given by the multiplication of their individual probabilities.

FCBF works well in removing irrelevant and redundant features from high-dimensional data, yet it lacks the capability to consider how the features work as a whole [66].

While FCBF has an entropy-based filtering criteria, other approaches focus on feature variance, independency and other metrics to establish a ranking, such that $J(x_{i_1}) \leq J(x_{i_2}) \leq \cdots \leq J(x_{i_N})$, where J refers to the criteria function, that may evidence the features to be filtered out.

4.2.2 Wrapper Approaches

A wrapper approach, receives it's name for wrapping the learning algorithm to assess it's iterations performance in order to optimize the set of features used. It becomes clear that having the learning method as a subroutine results in a heavier process than those offered by filter approaches and makes this approach less generalizable. However, the iterations provide an evaluation of the overall performance of the feature set, which tends to outperform the results of simply ranking and filtering features.

The algorithms following this approach assess the quality of the model produced by the learning procedure associated to each candidate feature subset. Thus, the learning procedure, itself, is an abstraction inside the wrapper approach with some generalization capabilities. The evaluation function applied to the model, usually estimates classification accuracy, although other metrics can be used. The difference between wrapper approaches, however, doesn't only focus on the metric used, but the search method as it would, easily, be overwhelming to evaluate every possible combination of features in our data.

Considering a set with N features, every possible combination would result in $2^N - 1$ possible subsets. A simple solution to reduce the number of considered subsets is to perform sequential search. The most simple greedy search algorithm is sequential forward search (SFS). It is implemented by starting without features and adding them one by one. At each iteration the performance is estimated to assess whether the feature improved or reduced the classification capability. Likewise a sequential backward search, starts with all features and iterates removing each feature to find an optimal subset.

A paper on feature selection for unsupervised learning [67] makes an attempt to apply a wrapper approach to the problem of unsupervised classification. They apply feature subset selection using expectation-maximization clustering algorithm to a wrapper framework as illustrated in Figure 4.2.

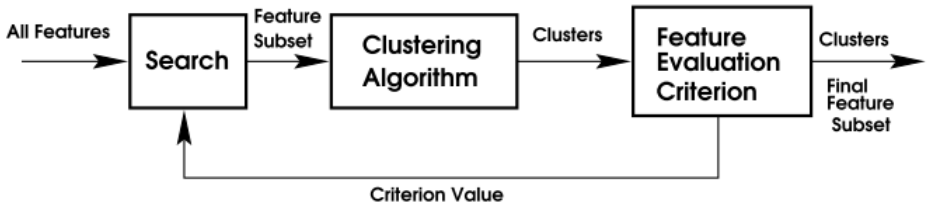


Figure 4.2: Flow diagram of wrapper approach for unsupervised learning [67, pp. 848]

The approach uses SFS as the search model and a clustering algorithm with a maximum likelihood (ML) criterion, attempting to produce clusters with Gaussian distributions, as the learning method. Subset selection attends to a scatter separability criterion, which maximizes between cluster and minimizes within cluster distances and to a ML criterion such as the learning method. The paper refers that, although a Gaussian distribution was used, the procedure could be extended to any other distribution.

Some of the feature selection methods considered in EHG related studies include:

- Single-feature classifier to determine individual discriminatory capacity and Sequential forward feature selection algorithm for determination of best combination of features [39].
- Sequential forward feature selection (SFFS) [34];
- Sequential forward feature selection (SFFS) assessing Receiver Operator Curve (ROC) of Holdout Cross-Validation [26];
- particle swarm optimization (PSO) [29];
- binary particle swarm optimization (BPSO), Jeffrey divergence distance between the two classes histograms and SFFS [15].

4.3 Defined Features

4.3.1 Introduction

In the last section (4.1) feature extraction was broken down into EHG related studies, where features are considered characteristics that attempt to describe relevant information of the signals, and general feature extraction methods, with a special focus on PCA.

After a revision of used EHG Features, general feature extraction methods were addressed, and in particular the PCA algorithm, which creates a lower dimensional representation of the data by minimizing feature correlation.

It was then revised in subsection 4.2 some methods of feature selection to reduce the initially extracted features to a more relevant subset.

A paper on feature selection for unsupervised leaning [67] describes the importance of eliminate redundant, irrelevant and misguiding features that may be initially considered and to decrease classification complexity. Unsupervised learning assumes no labels, therefore feature selection is domain-dependent, meaning the designer has to define what is relevant in a way to find the smallest feature sub-set that best defines "interesting natural"groups.

The objective of feature selection and extraction is to provide the best possible characterization of our data in a way that

- highlights classes through choosing characteristics with small intraclass variation and large interclass distance [68].
- Selects a number of features with a good compromise between providing most information and avoiding over-specification causing loss of generalization.
- Weights well the characteristics of a signal, meaning that it shouldn't have redundant features that define a characteristic to have overwhelming influence.
- Define well the domain of the data, with information representative of the different inputs our classifier may encounter.

Dimensionality reduction is, thus, a process of compressing information either from the measurements or a set of features to conciliate amount of information and it's relevance.

4.3.2 Feature Sets

The lack of knowledge about the characteristics of different contractions and about measurement's relevant information makes feature extraction the most difficult task in classifying such signals. Furthermore, the signal's dimensions and their non-linear characteristics promise high computational requirements for this task.

As to study what is relevant when studying EHG records, four sets of initial features are considered at this point.

1. Contraction segments in the time domain. Each of the contractions is then interpolated to obtain equally sized contractions and PCA was applied to define a linear transformation that best represents their differences.

2. Spectral representation of each contraction segment. Given that these already presented the same size, it was only applied the PCA procedure.

3. A number of linear, spectral and non-linear features referred in 4.1 that were extracted from the contraction segments and respective spectral representations.

4. Result of applying PCA to the latter set of features.

These sets are considered in an attempt to understand what features may hold more information in the study of contraction segments of EHG records.

The first and second sets of features are fairly simple to obtain. As we already have both the contractions and their spectral representations, it was only applied the PCA algorithm to attain a more concise representation.

The third set considered the following features:

- **Linear Features**: relative amplitude, kurtosis, standard deviation, autocorrelation first-zero crossing, log detector, maximum fractal length, mean absolute value, root mean square, simple square integral and average amplitude length;

- **Spectral Features**: contraction power, energy window, max power frequency, mean frequency, median frequency, deciles median frequencies, power ratio

- **Non Linear**: approximate entropy, correlation dimension, lyapunov exponent, sample entropy, time reversibility

The decision to use this many features, inclusively attempting the use of non-linear features, which have been described so far as inaccurate for such small signals, was to evaluate these characteristics at the feature selection level.

As such, a wrapper approach will be used that will iterate classifications to optimize the subset of features and a few criteria will be tested in this approach. Given the characteristics of the wrapper approach it will be described, when specifying the classification model.

CHAPTER 5

Machine Learning

5.1 Introduction

Characterized as the *Information Age*, today's world economy and science evolution highly depend on the acquisition and interpretation of data. The increasing computerization of industries and applications of computer-based approaches to daily-lives tools, pursued by solutions such as *Internet of Things* (IoT), sets information as one of the most crucial currencies to prosper at any market.

Data collected from sensors, software, forms, etc. has attained dimensions and complexity not interpretable by humans, and adding cost, coherence, timing, automation and many other benefits creates a migration of data analyzing to other machines and devices [64].

The classification problem requires a training set which should have representative information of the cases domain. The training set comprises input data to classify (causes, observations, records) and may have output data that states the class or response to the input data. Sets of pairs input-output allow a supervised classification, which consists of feeding the classifier the training set in order to adjust the classifier's processes to find the desired output for any specific input. The classifier should be implemented and trained in a way that the given examples by the training set will allow an accurate classification of new data. In a case where there is incomplete knowledge about the outputs of each training set data, semi-supervised learning approaches should be considered as they allow the best use of data in such situations (not discarding input data that does not offer the full set and using output data information). When there is no knowledge of the outputs, it comes down to unsupervised learning algorithms. This particular approach aims at finding a new representation or explanation of the observed data [64].

The purpose of this thesis can only be achieved with unsupervised classification as there is no knowledge about the classes the contractions considered should be placed in. In the case presented, a correspondent contraction class to each of the attained contractions would provide ground to perform a supervised classification and specify to the classifier how it should behave. That being said, correctly influencing the classifier depends on providing discriminative features of the contractions, which can accentuate their differences to provide

robust classes.

5.2 Unsupervised Learning

Unsupervised learning is defined as a process of finding patterns that distributes the input data into classes. These kinds of solutions are implemented in problems where there is no knowledge of what the outputted classes should be. Therefore, a number of algorithms are developed to partition the data in a way that makes sense, in accordance to some criteria.

The effort to bring meaning to the data and extracting features that would best characterize it, all sum up to the point where it is fed to the classification process. These are meant to point out the characteristics it should consider, just as normalization and redundant features removal is meant to distribute evenly the weights among characteristics, removing outliers avoids overwhelming influences from noise, data balancing attempts to equalize class weight and many more problem-dependent considerations can be taken into account before even getting here.

The classification initially undergoes a training phase where the patterns are identified or the behavior to given inputs is adjusted. The training phase is performed using a training set that should represent well the different classes, along the different values that may appear in the context that is being analyzed. After the training phase, usually comes a validation phase in which the generalization capabilities of the classifier are assessed. This phase will feed a different set denominated validation data, which tests how the classifier performs when dealing with new values. As such, the classification process has to deal with a compromise between accuracy of classification and generalization.

5.2.1 Data Clustering

Regarding an unsupervised problem with no knowledge of class probabilities, one approach that comes to mind is to perceive how the data is positioned geometrically, under the assumption that the samples may form clouds representative of a distribution characteristic of a class. Data clustering algorithms attempt to assess this patterns, to form groups of data, denominated clusters, with strong internal similarities and, preferably, strong external dissimilarities [69].

A first consideration while designing a clustering procedure is to define how the similarity between data samples is measured. This is a core consideration for clustering algorithms as it defines how the samples should relate to one another. It also depends on how the features considered are distributed. An obvious similarity measure can be given by a distance function as those defined by the general Minkowski metric [69] formulated as

$$d(x, x') = \left(\sum_{k=1}^{N} |x_k - x'_k|^q \right)^{\frac{1}{q}} \quad , \quad q \geq 1 \tag{5.1}$$

where the parameter q may be adjusted to define derived distance functions. The most used among these are listed in Table 5.1.

parameter q	**Metric**	**Description**
$q = 1$	Manhattan (or city block) distance	Measures the minimum gridlike (horizontal and vertical) distance from x_k to x'_k
$q = 2$	Euclidean distance	Measures the distance of the straight line from x_k to x'_k
$q = \infty$	Chebyshev distance	Measures the maximum absolute distance in one dimension of two N dimensional points

Table 5.1: General Minkowski Metric Derivations

As an example, euclidean distance may seem a reasonable choice. However, Euclidean distance may not be able to perceive the real distribution of the data, and it is highly sensible to scaling. Thus, the decision also requires an understanding of how our data can be represented. Normalizing the data may make euclidean distance a reasonable choice, but it might also reduce classes differences[69, ch. 10 pp. 27-28]. An example of when normalization may affect cluster separation is illustrated in figure 5.1.

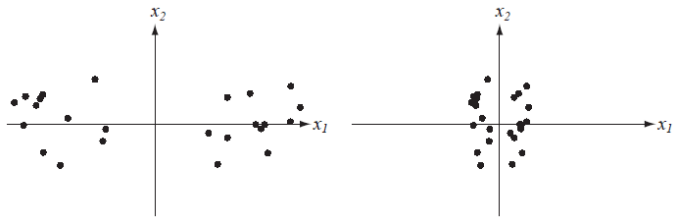


Figure 5.1: Example of how normalization may reduce cluster separation [69]

Other than distance functions a similarity measure may also be defined by a similarity function $s(x', x)$. For a case where the angle between vectors is a meaningful measure Richard O.Dudat et. al [69, ch. 10 pp. 27] describes the use of a cosine distance of the angle between x and x' as the similarity function given by

$$s(x, x') = \frac{x^T x'}{\|x\| \, \|x'\|}$$

where the superscript T refers to the transpose and $\|x\|$ refers to the Euclidean norm or length of the vector defined as

$$\|x\| = \sqrt{x^T x} \tag{5.2}$$

Defining a good similarity measure is the first challenge when designing a cluster algorithm and as afore mentioned it is a problem-dependent choice for which the referred measures may not be the best choice or even applicable. An example is for categoric binary data, to which Richard O. Duda et. al [69, ch. 10 pp. 28] refers two measures that asses the fraction of attributes shared and the ratio of the number of shared attributes to the

number possessed. Jasmine Irani et. al [70] describes the Jaccard distance given by the ratio of data intersection and union of the data sets, and many other similarity functions may be applied.

Another cluster designing specification is the criterion function to be optimized. This is the criteria for what defines a good partition of the data or, in other words, what is it trying to achieve.

The most intuitive criterion is sum-of-squared-error (SSE) minimization. It simply tries to create clusters so that member points are closer to the cluster centers, denominated centroids. As such, for a clustering iteration the centroids may be given by the mean of the points belonging to that cluster at that iteration. Let D be a set of n samples $x_1, \ldots, x_n$, which are being assigned at each iteration to one of c clusters by some similarity measure, so that n_i samples are assigned to the cluster D_i. Then the centroid m_i is defined as [69, ch. 10 pp. 29]

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x \tag{5.3}$$

and the SSE criterion measures the clustering quality as

$$J_e = \sum_{i=1}^{c} \sum_{x \in D_i} \|x - m_i\|^2$$

where $\|x - m_i\|$ refers to an euclidean norm defined as in 5.2.

The SSE criteria minimizes cluster scatter by evaluating the distance of the points to their respective cluster center using an euclidean distance. This works well when clusters form compact clouds well separated from each other. A problem with this criterion may persist in the fact that the number of samples in a cluster elevates the criterion value, which may lead to splitting large clusters that are over-represented in the training set [69, pp. 30]. Later problem is worsened by the existence of outliers which alone can already have big influences over this criterion.

Richard O. Duda et. al [69, ch. 10 pp. 30-31] demonstrates the derivation of a generalized criteria, which allows applying different similarity functions $s(x, x')$, from the SSE criteria. As described, it removes the mean vectors from SSE and manipulates the equation to obtain

$$J_e = \frac{1}{2} \sum_{i=1}^{c} n_i \bar{s}_i \tag{5.4}$$

where

$$\bar{s}_i = \frac{1}{n_i^2} \sum_{x \in D_i} \sum_{x' \in D_i} \|x - x'\|^2 \tag{5.5}$$

and the generalized criteria is, then, given by replacing the euclidean distance between x and x' by a similarity function $s(x, x')$, resulting in

$$\bar{s}_i = \frac{1}{n_i^2} \sum_{x \in D_i} \sum_{x' \in D_i} s(x, x'). \tag{5.6}$$

Through this simple manipulation, it is now possible to adapt the criteria to use a similarity measure that better fits the nature of the data.

Another criteria to consider is the scattering criteria. One, may define that the best clusters minimize within-cluster scatter, while maximizing between-cluster scatter. In accordance with the notation used while defining SSE criteria, the within-cluster scatter matrix is the sum of all cluster's scatter matrices, being defined as [69, ch. 10 pp. 31]

$$S_W = \sum_{i=1}^{c} \sum_{x \in D_i} (x - m_i)(x - m_i)^T \tag{5.7}$$

where m_i refers to the centroids already defined in 5.3. On the other hand, the between-cluster scatter is defined as [69, ch. 10 pp. 31]

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T \tag{5.8}$$

where m is the total mean vector of the data defined as [69, ch. 10 pp. 31]

$$m = \frac{1}{n} \sum_{x \in D} x = \frac{1}{n} \sum_{i=1}^{c} n_i m_i.$$

Lastly, the total scatter matrix is defined as [69, ch. 10 pp. 31]

$$S_T = \sum_{x \in D} (x - m)(x - m)^T = S_W + S_B \tag{5.9}$$

Given that S_T remains constant and is characteristic of the data set, it is implicit that minimizing S_W will also maximize S_B and vice-versa, which allows the criteria to focus on either one.

Given these definitions and notations, and the aim to minimize S_W and maximize S_B, Richard O. Duda et. al [69, ch. 10 pp. 32-33] make reference to a few additional criteria described in Table 5.2.

Criteria	**Description**
Trace Criteria	The trace of a matrix is the sum of it's diagonal elements, which in the case of a scattering matrix, that essentially corresponds to a feature covariance matrix, contains the feature variances. As such, when applied to the within-cluster scatter matrix, it minimizes the variance of features of the sames cluster. The within-cluster criterion is, thus, defined as $$tr\, S_W = \sum_{i=1}^{c} \sum_{x \in D_i} \|x - m_i\|^2 . \tag{5.10}$$ Which ends up being similar to the SSE criterion. Aiming to minimize $tr\, S_W$ is similar to maximizing between-scatter criteria $tr\, S_B$ and vice-versa.
Determinant Criteria	Captures information of the square of the scattering volume, being it proportional to the product of the variances in the directions of the principal axes [69, ch. 10 pp. 32]. The authors describe the criteria of minimizing the determinant $\|S_W\|$ to be applicable for nonsingular S_W, with results often similar to those of minimizing J_e and the advantage of being an approach that is invariant to scaling.

Criteria	Description
Invariant Criteria	Uses the fact that the eigenvalues $\lambda_1, \ldots, \lambda_N$ of $S_W^{-1} S_B$ are invariant under nonsingular linear transformations of the data, which is demonstrated in the book. The authors describe these as a measure of the ratio of between-cluster to within-cluster scatter in the direction of the eigenvectors and propose criteria such as the maximization of $$tr\ S_W^{-1} S_B = \sum_{i=1}^{N} \lambda_i$$ and the minimization of $$tr\ S_T^{-1} S_W = \sum_{i=1}^{N} \frac{1}{1+\lambda_i}$$ and $$\frac{|S_W|}{|S_T|} = \prod_{i=1}^{N} \frac{1}{1+\lambda_i}$$ being all of these criteria invariant to linear transformation.

Table 5.2: Cluster Criteria Functions

As of the point where a similarity measure and a criterion function have been chosen the clustering algorithm is ready to be implemented. The general implementation of a clustering algorithm will move the cluster centers along the iterations, while defining for each, what points belong to each class according to the similarity measure and the criterion function defines the stop condition. However, many algorithms exist that derive these general concepts to achieve different solutions.

5.2.2 Classification Methods used in Related Studies

Given the context of this work only unsupervised learning methods that require no prior knowledge of class probabilities will be considered.

Unsupervised classification: Fisher test [28] , wavelet networks.

- Fisher test to compare features, combined with K-means method [28]:

  1. Initialize center of first class with first vector.
  2. Take new vector and find it's nearest class center.
  3. Compare new vector to the classes by using Fisher method.
  4. If vector is identical to the class center given a false alarm probability adapt each class; otherwise create a new class.
  5. Go back to 2nd step.

5.3 Validation

Validation is the step of classification where the accuracy and generalization capability of the classifier are measured to validate the solution. A few important terms that are used in the validation context for supervised learning, where we know the class labels include:

- accuracy - Determined by the ratio of predictions the model correctly classified and total number of predictions. It measures the probability of the model correctly classifying data.

$$accuracy = \frac{correct\ predictions}{total\ number\ of\ predicitons}$$

- precision - Assesses how well a class is being classified, being the ration of correct predictions of a class by the number of values of that class.

$$precision = \frac{correct\ predictions\ of\ c_i}{number\ of\ samples\ that\ belong\ to\ c_i}$$

- recall - Determined by the ratio of predictions the model correctly classified in a class by the number of predictions that the model defined to belong to that class.

$$precision = \frac{correct\ predictions\ of\ c_i}{number\ of\ samples\ that\ belong\ to\ c_i}$$

- F-measure - Combines precision and recall defined as

$$F = (1 + \beta^2) \cdot \frac{precision \cdot recall}{\beta^2 \cdot precision + recall}$$

  where the weight β emphasize precision for $\beta < 1$ and emphasize recall for $\beta > 1$.

These represent the most common measures to validate a classification, for a labeled validation set, and emphasize what should be expected of a classifier.

In an unsupervised classification problem with no labels nor class probabilities, validation methods are more closely related to simply optimizing a criteria function as it was presented in 5.2.1. Although neither of the measures stated can be applied, considering relevant features and an appropriate criteria optimizing it, will provide clusters with a sort of validity. In terms of validating the results, one may consider how well characterized are the obtained clusters by their within-cluster coherence and between-cluster disparity, for example. However, if the results seem coherent and a good validation of the clustering procedure is obtained these are only defined by feature distribution with no reference to how well contextualized are the features to address our problem.

5.4 Classification Model For EHG Data

In section 4.1 it was defined the use of four sets of features. Here, it will be discussed what was attempted for each of those sets. As mentioned the four sets include:

- Contractions in the time domain. Each of the contractions is then interpolated to obtain equally sized contractions and PCA was applied to define a linear transformation that best represents their differences.
- Spectral representations of contractions. Given that these already presented the same size, it was only applied the PCA procedure.
- A number of linear, spectral and non-linear features referred in 4.1 that were extracted from the contraction segments and respective spectral representations.
- Result of applying PCA to the later set of features.

Each of these sets has the purpose of studying specific characteristics of the signals, and the classification procedure for each will be described in the next subsections.

5.4.1 PCA and K-means Combination

The same classification process was applied to the first two and last feature sets defined in 4.3.2. As aforementioned, the first set of features is composed of the detected contractions, which were re-sampled through interpolation to attain signals of the same size. This procedure does not change the shape of the contraction, rather rescales the sampling frequency of each contraction to a relative scale.

The second set of features didn't require a re-sampling procedure as the spectral representations of contractions already have the same size, being in function of frequency defined to range from 0Hz to 2Hz within an equal scale for all contractions.

The fourth set was described in 4.3.2, as having multiple linear time and frequency domain, and non-linear features. This set is, thus, formed of some computed values from the contractions and their spectral representations.

As introduced in subsection 4.3.2, PCA is an algorithm that creates a new coordinate system which combines the initial feature set to better represent it according to a decorrelation optimization criteria. It, not only allows, eliminating dimensions, but also the new coordinate system should create a more intuitive representation of the data.

The feature sets are, first, applied a linear transformation by the PCA algorithm and then are clustered by a simple implementation of a k-means algorithm. Considering that it is intended the separation of data into k clusters, with Euclidean distance as the metric function. The clustering algorithm is defined by the following steps:

1. Initialize k random centroids;
2. Define cluster for each signal based on which centroid is closer to the signal according to the metric function;
3. Calculate new Centroids. Each of the k centroids are a vector containing the mean of each feature from it's respective cluster.

4. If the centroids haven't moved an optimal point has been reached, otherwise repeat from step 2.

This procedure is repeated a few times to minimize sum-of-squared-error criteria. The repetition of the clustering algorithm a few times allows optimizing the results that are influenced by the fact that initial centroids are being chosen randomly, which may lead to different local optima.

5.4.2 Wrapped Clustering Model

The classification process applied to the third set of features, which were described in subsection 4.3.2, uses a more general clustering algorithm, wrapped in a feature selection procedure, which calls it successively to test out feature subsets.

The wrapped classifier has two phases, which perform a version of a sequential backward search (SBS) and of a sequential forward search (SFS), respectively. The model was designed to study the features as well as the clustering criteria and metrics. In this model it is hypothesized, that the optimization of accurately defining the one class that is known may be regard useful, in terms of identifying the underlying physiological properties of these signals. As such, the model implements an unsupervised classification with a supervised optimization.

It is important to note that optimizing the correct classification of one class, holds no guarantees in providing a beneficial effect on the others. However, it's still considered useful by the author with regard to the analyzes of the feature sets and the clustering methods.

The model is described by the following steps:

1. The algorithm starts considering all features;
2. Apply clustering procedure;
3. Remove the worse feature according to a given criteria;
4. Apply clustering procedure to new subset of features;
5. Assess whether the classification accuracy has decreased:
    - If it has, add the removed feature as an elite (won't be removed anymore) and go back to step 2;
    - If it hasn't, check whether all considered features are elite. If they aren't go back to step 3, otherwise continue to step 6;
6. Add a non-elite feature;
7. Apply clustering procedure;
8. Assess whether the classification accuracy has decreased:

- If it has, discard added feature;
- If it hasn't, continue;

9. If there are still non-elite feature that haven't been tested again go back to step 6.

Figure 5.2 represents an illustration of the model, where the green and blue areas refer to SBS and SFS parts, respectively. The model was designed to have both the removal and recuperation of features parts, as only the removal would leave too few features, and although those could distinguish well the fetal movements, they may not characterize well the differences between the types of contractions. As such, features that would not have a negative effect in fetal movement classification were recuperated. Furthermore, the clustering procedure does not always have good results. Even though it is performed five times at each iteration to attain better clusters according to the criteria, it may misrepresent features at some iteration, which then get another chance in the SFS.

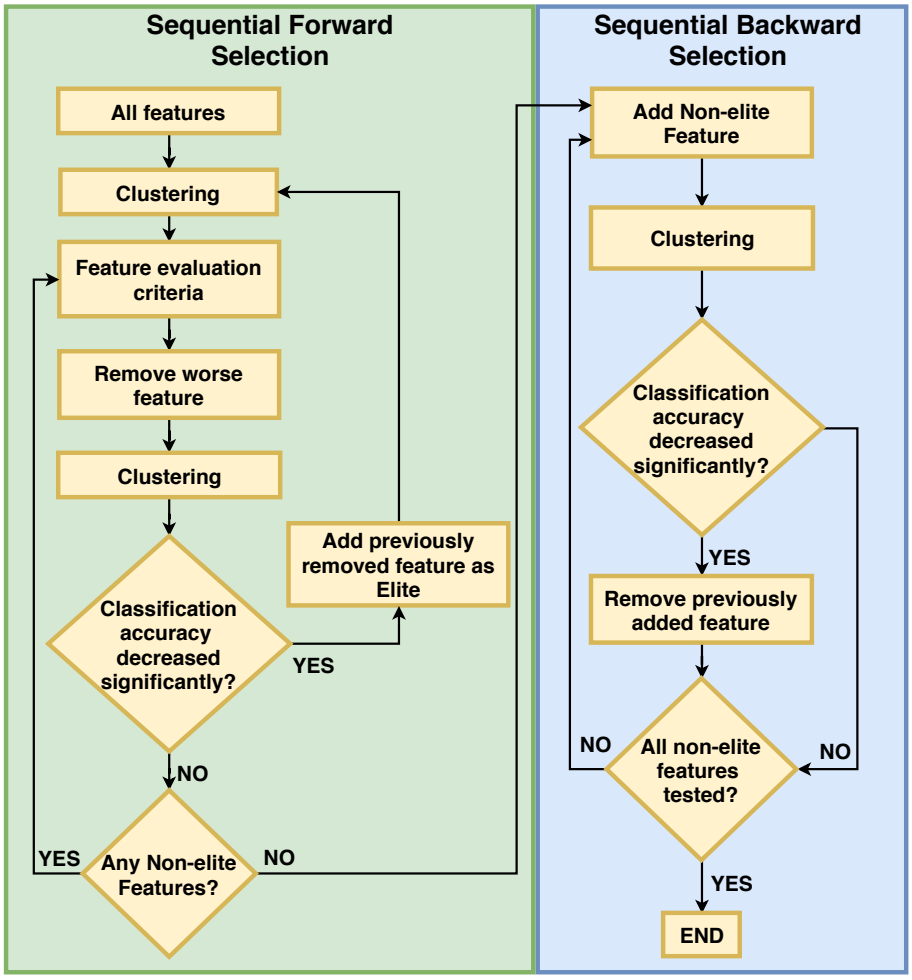


Figure 5.2: Developed Wrapper Model for feature selection. Green and blue areas refer to SBS and SFS parts, respectively. The diamond boxes represent conditions, whereas rectangular boxes represent stages or procedures.

The criteria to remove or add a feature considers whether there has been a loss in fetal movements clustered together, along whether contractions have been wrongly classified as fetal movements. Thus, it maximizes the ratio of fetal movements well classified and minimizes the ratio of contractions classified as fetal movements.

CHAPTER 6

Project Implementation

6.1 Introduction

In this chapter the project's implementation is described in terms of the programs created to allow testing and evaluating the described approaches. Given the goal of creating an unsupervised method to distinguish different types of contractions, two programs were developed. The first program performs signal processing and classification procedures, while the second allows signals, contractile segments and classification results inspection. These programs will be covered in the next two sections to describe what they do and the intention behind their development and approaches.s

6.2 Signal Processing and Classification Program

The signal processing and classification program was developed in Matlab and implements the procedures described in this thesis. The purpose of this program is to perform all procedures necessary from the provided raw signals to process them extract their features classify them.

Going through the procedure it will be bronken down into two parts. The first consisting on processes applied to our signals to build a database of processed signals ready to be classified. The second includes PCA and wrapper models execution.

6.2.1 Signal Processing

The first part of our program is represented by a flow diagram in Figure 6.1, by means of functions called and a brief explanation of their execution, as well as the evolution of our database, along procedures.

A few local procedures of the main function are not represented in the diagram as they are mostly organizational procedures, but will be described along this section. Further explanation of the diagram will refer to the green blocks as main execution blocks.

The first step of our procedure creates a profile for every patient and performs the first preprocessing procedures described in the thesis up to bipolar signal computation.

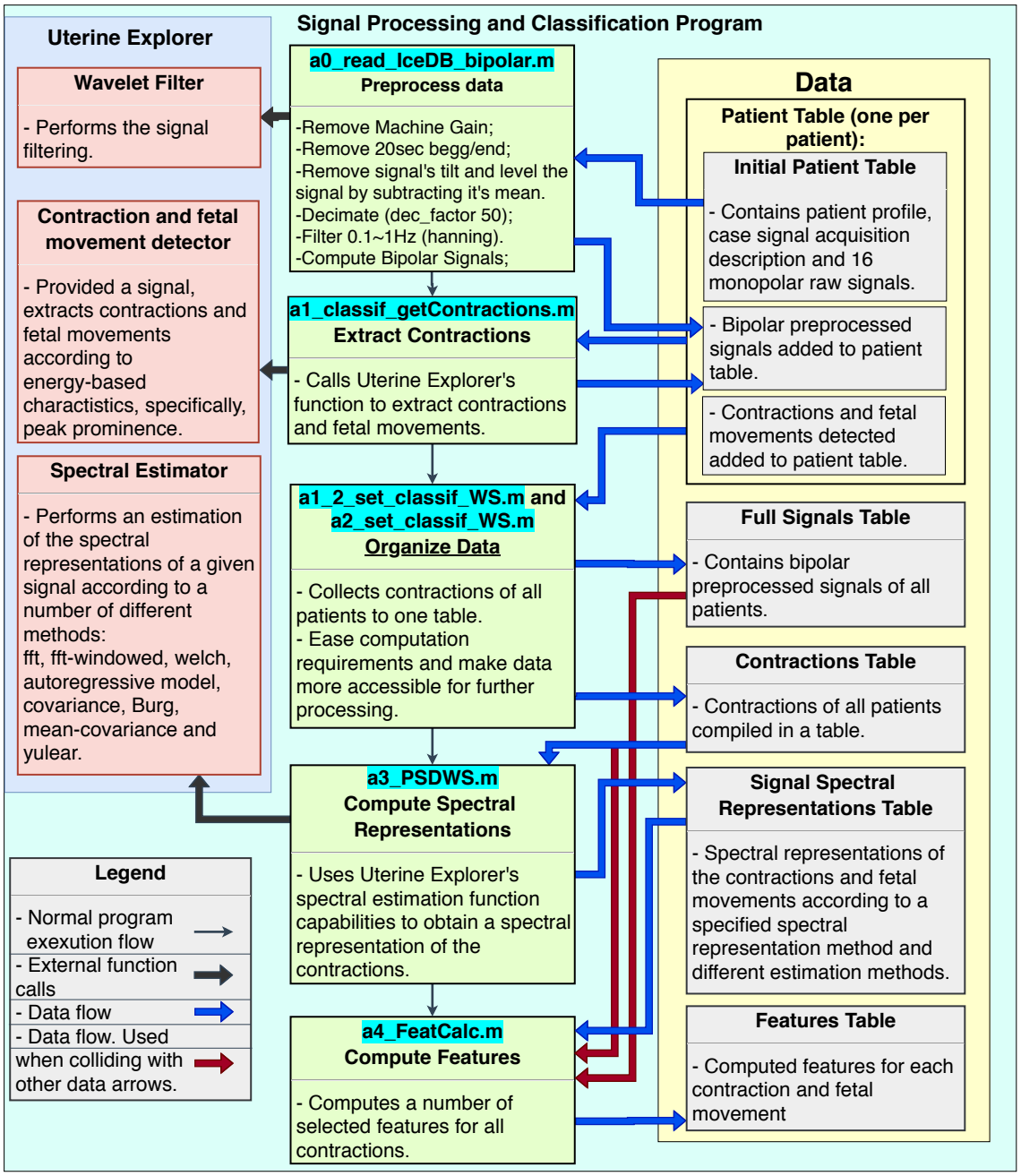


Figure 6.1: Flow diagram summarizing signal processing procedures. Pink boxes refer to functions external to the developed program, which belong to UEX project scripts. Green boxes refer to the scripts developed that implement the main flow of the program. Scripts names are are shown with a cyan background. The beige boxes represent data. The diagram provides a legend for arrows in the grey box.

It's input consists of two files for each patient, one provides the 16 monopolar signals and another with the patient's profile. These files can be found in PhysioNet's database [3], although signal record files have to be converted from WFDB to matlab's '.mat' format prior to this procedure.

First block's output will update each file folder to a structure represented in Figure 6.2

19x1 table

	1 Ice001_L_1of1m_decimated
1 Data	*16x1840 double*
2 Filtered Data	*10x1840 double*
3 Gestational Weeks	39.4286
4 Record Type	'Labour'
5 Delivery Week	39.4286
6 Type of Delivery	'Vaginal'
7 Oxytocin in Labour	'No'
8 Epidural	'No'
9 IBM before	23.3000
10 IBM at recording	27.6000
11 Age	'31'
12 Gravidity	'3'
13 Parity	'2'
14 Previous Caesarean	'No'
15 Placental Position	'Fundus'
16 Comments for recording	*1x2 cell*
17 Comments for delivery	*1x1 cell*
18 Frequency	4
19 Filter wavelength	[0.1000,1]

Figure 6.2: Example output of script execution. Provides a structure with patient case profile and processed signal.

In the table presented the first field contains the signals to which all processing procedures other than filtering and computation of bipolar signals were applied. The second field contains the fully processed signals, which will be used for further processing. All other fields define patient's case profile with pregnancy and the mother's relevant information provided by Physionet's database [3].

The second main execution block represented in the flow diagram refers to the extraction of uterine events, which calls an UEX function, which outputs segments which were considered contractions. These are saved in a cell of cells and appended to the structure in 6.2 for further use.

For further processing simplification, the third execution block defines a new table joining all patients, which is represented in Figure 6.3.

Figure 6.3 shows a table with a patient case per row displaying their profile and the uterine events extracted. The zoomed in cell of cells in *contractions* column shows 10 rows, referring to the bipolar signal from which the contractile segment was extracted, and 5 columns referring to the energy estimation method used. These correspond to wavelet, Teager-kaiser, RMS, RMS squared and Hilbert-Huang energy estimations, respectively. The same logic applies to *fetalMov* column, which contains the extracted fetal movements, as well as for *fetalMIdx* and *contracIdx*, containing initial and final sample positions of the event extracted in the respective full record signal.

Figure 6.3: Example output of a1_2_set_Classif_WS.m script execution. Provides a structure with patient case profile and uterine event extracted.

The same block of the diagram, refers a second script that creates a table where each row corresponds to a contractions 6.4. This table optimizes iterating over each contraction and removes focus over patient cases. Each contractions location can be traced back by identification fields represented in the first 7 columns, as well as the *CS_Idx* column, that provides the same information as the earlier described *contracIdx* column.

This structure was chosen to allow iterating over each contraction to ease feature extraction and analysis procedure as well classification.

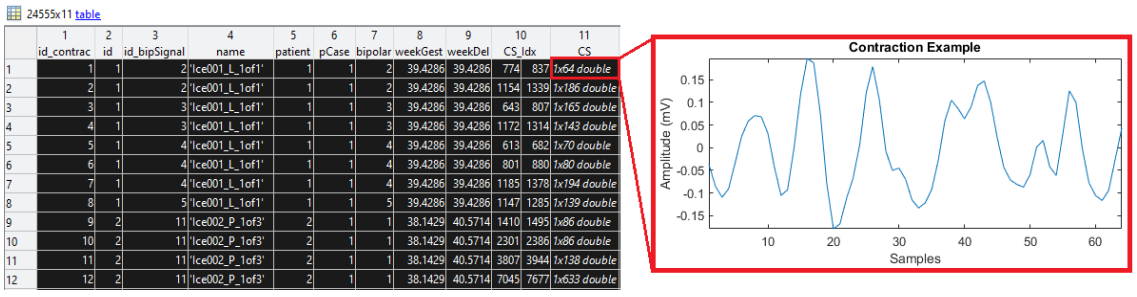
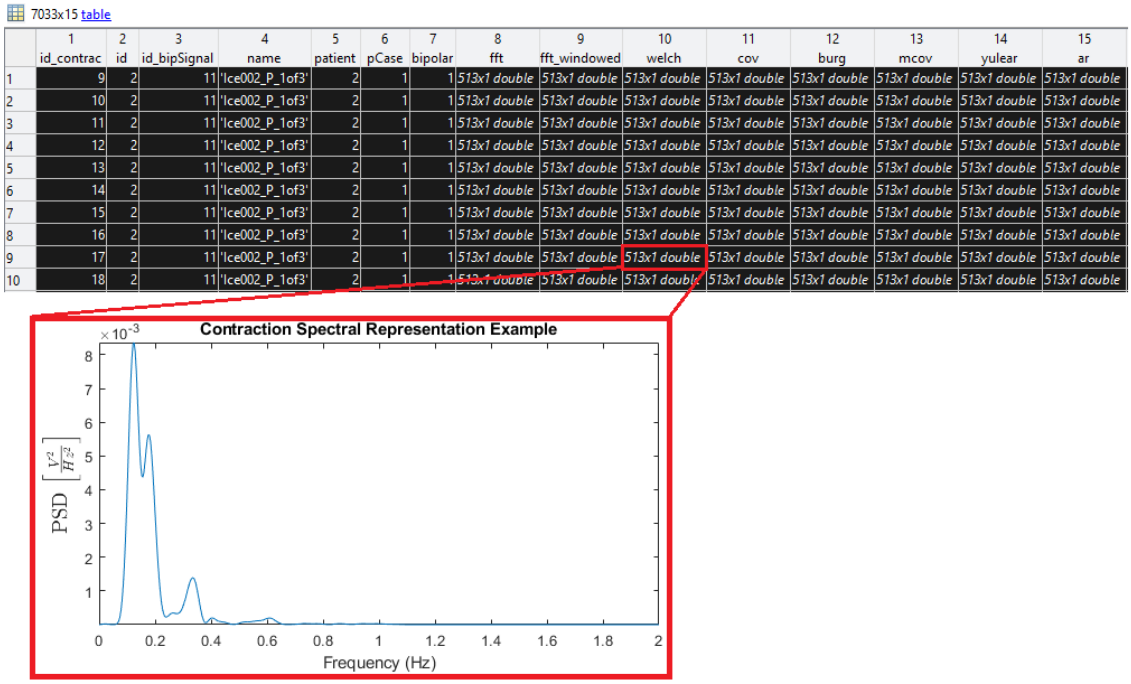


Figure 6.4: Example output of a2_set_Classif_WS.m script execution. Provides a structure with each contraction extracted per row and information of where it was extracted from along with respective patient case profile.

On thing to be noted is that fetal movements are no longer in the output table represented in 6.4 as they were saved in a different table with the same structure.

Third main execution block of the diagram runs *a3_PSDWS.m* script which performs spectral representation estimations of the contractions. To execute this task an UEX function is called which outputs the mentioned estimations by a few different methods. The third block's output is represented in Figure 6.5

The table presented in Figure 6.5 provides indexing information to it back to earlier



	1 id_contrac	2 id	3 id_bipSignal	4 name	5 patient	6 pCase	7 bipolar	8 fft	9 fft_windowed	10 welch	11 cov	12 burg	13 mcov	14 yulear	15 ar
1	9	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
2	10	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
3	11	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
4	12	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
5	13	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
6	14	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
7	15	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
8	16	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
9	17	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double
10	18	2	11	'Ice002_P_1of3'	2	1	1	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double	513x1 double

Figure 6.5: Example output of a3_PWDWS.m script execution. Provides a structure with each contraction's spectral representations by different methods computed in a UEX function execution.

tables, as well as spectral estimations provided by the fast Fourier transform, windowed fast Fourier transformed, Welch, covariance, Burg, modified covariance, Yulear and Autoregressive model methods.

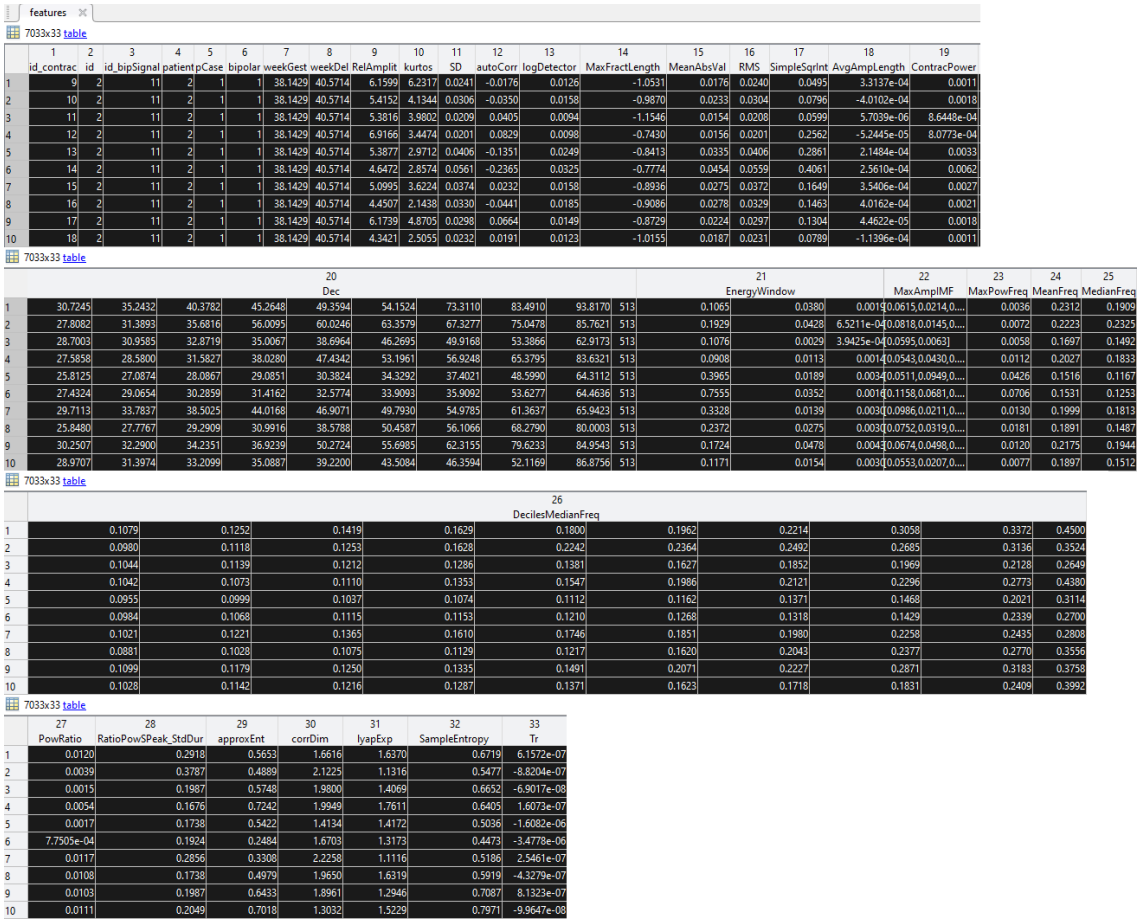
The last main execution block of the diagram 6.1, considers the script that computes the considered features. This script collects both time and frequecy domain representations of the signal to compute all the feature that will be evaluated by the wrapper model. An example of the output provided by the script is represented in Figure 6.6

Again, there is the necessity to provide a tracing back capability to the table, which is provided by the first six columns. Seventh and eighth columns are not used by the wrapper model either. Their addition to the table intents allowing the study of correlations between obtained features and both gestational and delivery weeks.

According to column order features shows, include:

- Time linear features: Relative amplitude, kurtosis, standard deviation, autocorrelation, log detector, max fractal length, mean absolute value, root mean square, simple square root and average amplitude length;

- Frequency linear features: Contraction power, deciles limits sample, energy distribution over three defined windows, Maximum Amplitude of intrisic mode function provided by Hilbert-Huang decomposition, power of peak frequency, mean and median frequencies, median frequency of each decile, ratio between peak frequency and standard deviation of burst duration;



features

7033x33 table

	1 id_contrac	2 id	3 id_bipSignal	4 patient	5 pCase	6 bipolar	7 weekGest	8 weekDel	9 RelAmplit	10 kurtos	11 SD	12 autoCorr	13 logDetector	14 MaxFractLength	15 MeanAbsVal	16 RMS	17 SimpleSqrInt	18 AvgAmpLength	19 ContracPower
1	9	2	11	2	1	1	38.1429	40.5714	6.1599	6.2317	0.0241	-0.0176	0.0126	-1.0531	0.0176	0.0240	0.0495	3.3137e-04	0.0011
2	10	2	11	2	1	1	38.1429	40.5714	5.4152	4.1344	0.0306	-0.0350	0.0158	-0.9870	0.0233	0.0304	0.0796	-4.0102e-04	0.0018
3	11	2	11	2	1	1	38.1429	40.5714	5.3816	3.9802	0.0209	0.0405	0.0094	-1.1546	0.0154	0.0208	0.0599	5.7039e-06	8.6448e-04
4	12	2	11	2	1	1	38.1429	40.5714	6.9166	3.4474	0.0201	0.0829	0.0098	-0.7430	0.0156	0.0201	0.2562	-5.2445e-05	8.0773e-04
5	13	2	11	2	1	1	38.1429	40.5714	5.3877	2.9712	0.0406	-0.1351	0.0249	-0.8413	0.0335	0.0406	0.2861	2.1484e-04	0.0033
6	14	2	11	2	1	1	38.1429	40.5714	4.6472	2.8574	0.0561	-0.2365	0.0325	-0.7774	0.0454	0.0559	0.4061	2.5610e-04	0.0062
7	15	2	11	2	1	1	38.1429	40.5714	5.0995	3.6224	0.0374	0.0232	0.0158	-0.8936	0.0275	0.0372	0.1649	3.5406e-04	0.0027
8	16	2	11	2	1	1	38.1429	40.5714	4.4507	2.1438	0.0330	-0.0441	0.0185	-0.9086	0.0278	0.0329	0.1463	4.0162e-04	0.0021
9	17	2	11	2	1	1	38.1429	40.5714	6.1739	4.8705	0.0298	0.0664	0.0149	-0.8729	0.0224	0.0297	0.1304	4.4622e-05	0.0018
10	18	2	11	2	1	1	38.1429	40.5714	4.3421	2.5055	0.0232	0.0191	0.0123	-1.0155	0.0187	0.0231	0.0789	-1.1396e-04	0.0011

7033x33 table

	20 Dec									21 EnergyWindow		22 MaxAmpIMF	23 MaxPowFreq	24 MeanFreq	25 MedianFreq
1	30.7245	35.2432	40.3782	45.2648	49.3594	54.1524	73.3110	83.4910	93.8170	513	0.1065	0.0380	0.0015	[0.0615,0.0214,0....	0.0036	0.2312	0.1909
2	27.8082	31.3893	35.6816	56.0095	60.0246	63.3579	67.3277	75.0478	85.7621	513	0.1929	0.0428	6.5211e-04	[0.0818,0.0145,0....	0.0072	0.2223	0.2325
3	28.7003	30.9585	32.8719	35.0067	38.6964	46.2695	49.9168	53.3866	62.9173	513	0.1076	0.0029	3.9425e-04	[0.0595,0.0063]	0.0058	0.1697	0.1492
4	27.5858	28.5800	31.5827	38.0280	47.4342	53.1961	56.9248	65.3795	83.6321	513	0.0908	0.0113	0.0014	[0.0543,0.0430,0....	0.0112	0.2027	0.1833
5	25.8125	27.0874	28.0867	29.0851	30.3824	34.3292	37.4021	48.5990	64.3112	513	0.3965	0.0189	0.0034	[0.0511,0.0949,0....	0.0426	0.1516	0.1167
6	27.4324	29.0654	30.2859	31.4162	32.5774	33.9093	35.9092	53.6277	64.4636	513	0.7555	0.0352	0.0016	[0.1158,0.0681,0....	0.0706	0.1531	0.1253
7	29.7113	33.7837	38.5025	44.0168	46.9071	49.7930	54.9785	61.3637	65.9423	513	0.3328	0.0139	0.0030	[0.0986,0.0211,0....	0.0130	0.1999	0.1813
8	25.8480	27.7767	29.2909	30.9916	38.5788	50.4587	56.1066	68.2790	80.0003	513	0.2372	0.0275	0.0030	[0.0752,0.0319,0....	0.0181	0.1891	0.1487
9	30.2507	32.2900	34.2351	36.9239	50.2724	55.6985	62.3155	79.6233	84.9543	513	0.1724	0.0478	0.0043	[0.0674,0.0498,0....	0.0120	0.2175	0.1944
10	28.9707	31.3974	33.2099	35.0887	39.2200	43.5084	46.3594	52.1169	86.8756	513	0.1171	0.0154	0.0030	[0.0553,0.0207,0....	0.0077	0.1897	0.1512

7033x33 table

	26 DecilesMedianFreq									
1	0.1079	0.1252	0.1419	0.1629	0.1800	0.1962	0.2214	0.3058	0.3372	0.4500
2	0.0980	0.1118	0.1253	0.1628	0.2242	0.2364	0.2492	0.2685	0.3136	0.3524
3	0.1044	0.1139	0.1212	0.1286	0.1381	0.1627	0.1852	0.1969	0.2128	0.2649
4	0.1042	0.1073	0.1110	0.1353	0.1547	0.1986	0.2121	0.2296	0.2773	0.4380
5	0.0955	0.0999	0.1037	0.1074	0.1112	0.1162	0.1371	0.1468	0.2021	0.3114
6	0.0984	0.1068	0.1115	0.1153	0.1210	0.1268	0.1318	0.1429	0.2339	0.2700
7	0.1021	0.1221	0.1365	0.1610	0.1746	0.1851	0.1980	0.2258	0.2435	0.2808
8	0.0881	0.1028	0.1075	0.1129	0.1217	0.1620	0.2043	0.2377	0.2770	0.3556
9	0.1099	0.1179	0.1250	0.1335	0.1491	0.2071	0.2227	0.2871	0.3183	0.3758
10	0.1028	0.1142	0.1216	0.1287	0.1371	0.1623	0.1718	0.1831	0.2409	0.3992

7033x33 table

	27 PowRatio	28 RatioPowSPeak_StdDur	29 approxEnt	30 corrDim	31 lyapExp	32 SampleEntropy	33 Tr
1	0.0120	0.2918	0.5653	1.6616	1.6370	0.6719	6.1572e-07
2	0.0039	0.3787	0.4889	2.1225	1.1316	0.5477	-8.8204e-07
3	0.0015	0.1987	0.5748	1.9800	1.4069	0.6652	-6.9017e-08
4	0.0054	0.1676	0.7242	1.9949	1.7611	0.6405	1.6073e-07
5	0.0017	0.1738	0.5422	1.4134	1.4172	0.5036	-1.6082e-06
6	7.7505e-04	0.1924	0.2484	1.6703	1.3173	0.4473	-3.4778e-06
7	0.0117	0.2856	0.3308	2.2258	1.1116	0.5186	2.5461e-07
8	0.0108	0.1738	0.4979	1.9650	1.6319	0.5919	-4.3279e-07
9	0.0103	0.1987	0.6433	1.8961	1.2946	0.7087	8.1323e-07
10	0.0111	0.2049	0.7018	1.3032	1.5229	0.7971	-9.9647e-08

Figure 6.6: Example output of a4_FeatCalc.m script execution. Aggregates computed values that attempt to elevate the uterine event's time, spectral and non-linear characteristics. Each row of computed values is associated to an uterine event.

- Non-linear features: Approximate entropy, correlation dimension, Lyapunov exponent, sample entropy and time reversibility.

From the aforementioned features, decile limits, energy across three defined windows, maximum amplitude of intrinsic mode function and ratio between peak frequency and standard deviation of burst were removed and were not used considered in the wrapper approach.

At this point, instead of signals represented in the time and frequency domain, what represents our data are single values computed to extract condensed information of the signal. Thus, this table's set data shows an attempts to more efficiently define time, frequency and non-linear characteristics of the signals.

An omitted step on the diagram is a local function to join fetal movements and contractions in the same structure, at which a column specifying whether or not the event is a fetal movement is added.

6.2.2 Feature Selection and Classification Procedures

The second part of the developed program aggregates feature selection and classification procedures. Figure 6.7 shows the flow diagram that may be interpreted as a continuation of that of Figure 6.1, including with regards to the arrows legend, which is not displayed in 6.7.

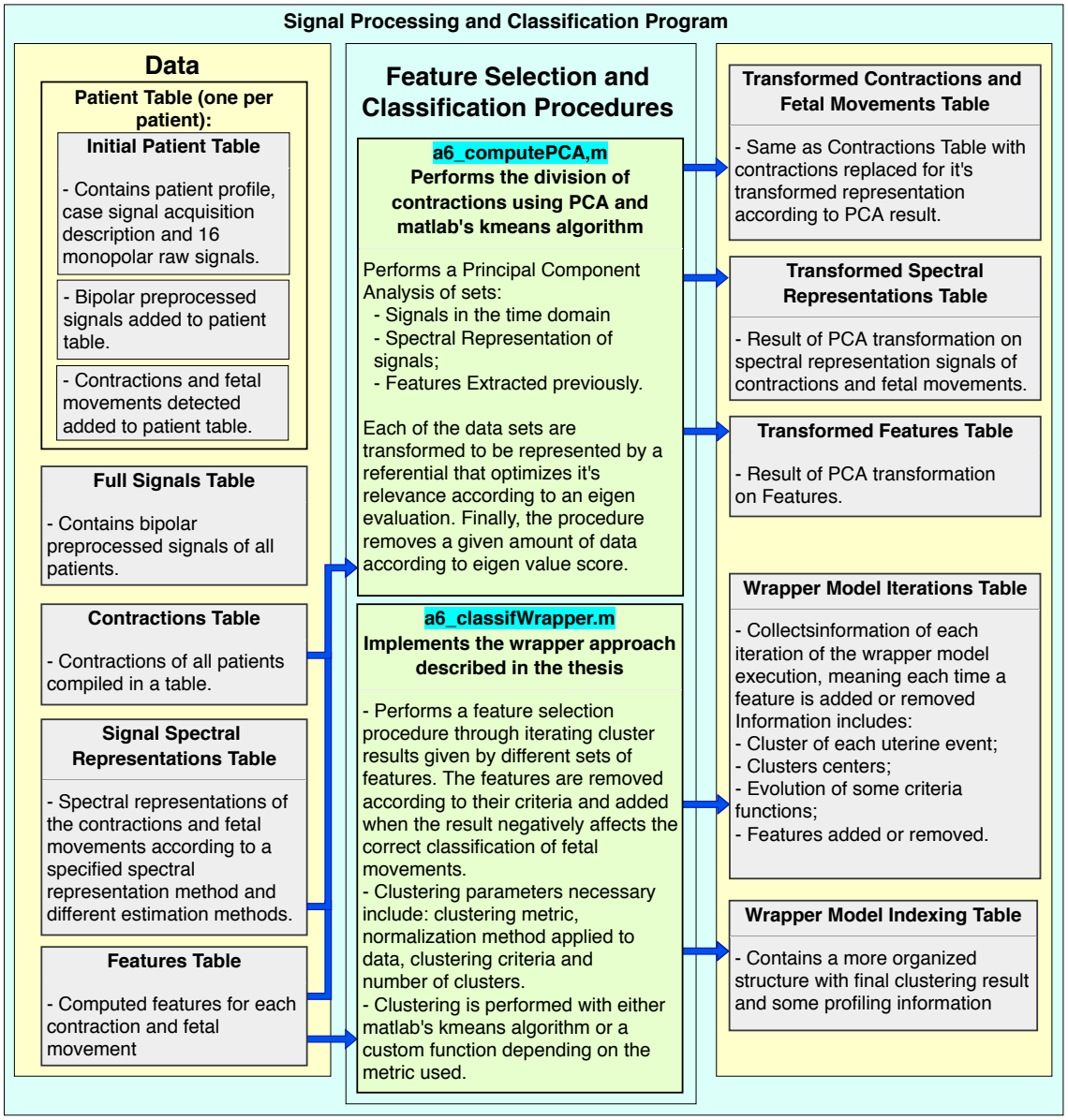


Figure 6.7: Flow diagram summarizing the developed procedures for feature selection and uterine events classification. Green boxes refer to the scripts developed that implement the main flow of the program. Script names are shown with a cyan background. The beige boxes represent data. The diagram provides a legend for arrows in the grey box.

The diagram 6.7 shows two main flow blocks, independent of each other. At the stage represented, all our feature sets are ready to be evaluated and applied to the classification procedures.

The first block presented addresses a principal component analysis applied to three different feature sets composed by time and frequency domain representations of the uterine events for the first two sets, and computed features as the third feature set. Output of PCA execution with regards to time representation set is present in figure 6.8.

13124x14 table

	1 type	2 id_contrac	3 id	4 id_bipSignal	5 name	6 patient	7 pCase	8 bipolar	9 weekGest	10 weekDel	11 CS_Idx		12 CS	13 PCArep	14 cluster
1	2	1	1	1	'Ice001_L_1of1'	1	1	1	39.4286	39.4286	60	107	1x512 double	20x1 double	3
2	2	2	1	1	'Ice001_L_1of1'	1	1	1	39.4286	39.4286	442	494	1x512 double	20x1 double	3
3	2	3	1	1	'Ice001_L_1of1'	1	1	1	39.4286	39.4286	974	1023	1x512 double	20x1 double	3
4	2	4	1	1	'Ice001_L_1of1'	1	1	1	39.4286	39.4286	1441	1499	1x512 double	20x1 double	3
5	2	5	1	1	'Ice001_L_1of1'	1	1	1	39.4286	39.4286	1519	1566	1x512 double	20x1 double	3
6	2	44	1	9	'Ice001_L_1of1'	1	1	9	39.4286	39.4286	218	258	1x512 double	20x1 double	2
7	2	45	1	9	'Ice001_L_1of1'	1	1	9	39.4286	39.4286	321	395	1x512 double	20x1 double	4
8	2	46	1	9	'Ice001_L_1of1'	1	1	9	39.4286	39.4286	450	512	1x512 double	20x1 double	3
9	2	47	1	9	'Ice001_L_1of1'	1	1	9	39.4286	39.4286	1109	1169	1x512 double	20x1 double	4
10	2	48	1	9	'Ice001_L_1of1'	1	1	9	39.4286	39.4286	1441	1490	1x512 double	20x1 double	2
11	2	49	1	9	'Ice001_L_1of1'	1	1	9	39.4286	39.4286	1721	1769	1x512 double	20x1 double	2

Figure 6.8: Example output of a6_computePCA.m script execution. This output exemplifies a file generated when applying PCA to the uterine event in the time domain. The first 10 rows provide identification and profile information, and in specific the first row specifies whether the row corresponds to a fetal movement. Column 12 has the contractile events considered in the PCA procedure. These all present the same size because each uterine event was interpolated to be represented in a 512 sample signal. Column 13 shows the resulting transformed and reduced result of PCA execution. According to the evolution of the eigen vector rating, the transformed signals were reduced to only account for 20 samples. The last column shows the cluster assigned to each signal.

While PCA over time representation set's output was provided Spectral representation and computed feature sets present a similar structure. Spectral representation set's output doesn't show the contractions but rather their spectral estimations and the respective reduced transformation, which for the case of wavelet contraction separator with welch's PSD estimation using a hanning window the obtained reduced transformation was composed of only 6 samples. Computed values set's output, in the other hand presents 32 columns of features that were transformed and reduced to 6 values of higher relevancy according to the eigen value criteria.

The application of PCA to the three mentioned sets intents on performing a comparison of our selected features set in the wrapper model to other possible approaches that only regard time or frequency characteristics relevant to separate the uterine events.

The second and last main flow block of the diagram is an abstraction of the wrapper model explained in 5.4.2 and is independent of PCA execution, being rather two different paths than an execution flow. Briefly, the model attempt a clustering procedure many times removing and adding back features to optimize the feature set according to some criteria. This procedure was developed to assess the relative lack of credibility associated to the validation of an unsupervised learning procedure. When a clustering solution is found it's validation is associated to scores found over how the features of a given

cluster's members are distributed with respect to distance, dispersion or other evaluating criteria. However, it does not assess how contextualized are the features with respect to the characterization and separation of the real classes. Thus, the wrapper model implemented offers a feature selection procedure under the hypothesis that accurately separating the one known class will better optimize our set of features. Furthermore, this hypothesis predicates on assuming that the selected features may hold better discriminative information from the uterine events.

The outputs of a wrapper model execution is shown in Figures 6.9 and 6.10.

69x17 table

	1 iter	2 type	3 clust	4 centr	5 Je	6 trSw	7 trSb	8 Jd	9 Jf
1	1	12981x1 do...	12981x1 do...	33x4 double	6.8552e+03	2.2622e+05	1.4917e+05	2.7252e+101	5.7747
2	2	12981x1 do...	12981x1 do...	33x4 double	6.8552e+03	2.2622e+05	1.4917e+05	2.7268e+101	5.7735
3	3	12981x1 do...	12981x1 do...	32x4 double	6.6663e+03	2.1332e+05	1.4916e+05	2.1362e+97	5.7710
4	4	12981x1 do...	12981x1 do...	31x4 double	6.5079e+03	2.0175e+05	1.4804e+05	2.9252e+93	5.7751
5	5	12981x1 do...	12981x1 do...	30x4 double	6.3446e+03	1.9034e+05	1.4804e+05	2.5886e+89	5.7717
6	6	12981x1 do...	12981x1 do...	29x4 double	6.1875e+03	1.7944e+05	1.4796e+05	2.4522e+85	5.7611
7	7	12981x1 do...	12981x1 do...	28x4 double	6.0447e+03	1.6925e+05	1.4517e+05	2.6821e+82	5.7978
8	8	12981x1 do...	12981x1 do...	27x4 double	5.9018e+03	1.5935e+05	1.4278e+05	1.1920e+79	5.7876
9	9	12981x1 do...	12981x1 do...	26x4 double	5.7749e+03	1.5015e+05	1.4053e+05	5.0521e+75	5.7633
10	10	12981x1 do...	12981x1 do...	25x4 double	5.5177e+03	1.3794e+05	1.3975e+05	1.2404e+72	5.7086

69x17 table

	10 Jf2	11 Jef	12 trSwf	13 trSbf	14 Jdf	15 Jff	16 Jf2f	17 idx
1	0.0415	33x1 double	33x1 double	33x1 double	33x1 double	33x1 double	33x1 double	'initial'
2	0.0415	33x1 double	33x1 double	33x1 double	33x1 double	33x1 double	33x1 double	'-corrDim'
3	0.0415	32x1 double	32x1 double	32x1 double	32x1 double	32x1 double	32x1 double	'-SampleEntropy'
4	0.0415	31x1 double	31x1 double	31x1 double	31x1 double	31x1 double	31x1 double	'-AvgAmpLength'
5	0.0415	30x1 double	30x1 double	30x1 double	30x1 double	30x1 double	30x1 double	'-autoCorr'
6	0.0416	29x1 double	29x1 double	29x1 double	29x1 double	29x1 double	29x1 double	'-DecilesMedianFreq10'
7	0.0412	28x1 double	28x1 double	28x1 double	28x1 double	28x1 double	28x1 double	'-kurtos'
8	0.0413	27x1 double	27x1 double	27x1 double	27x1 double	27x1 double	27x1 double	'-DecilesMedianFreq1'
9	0.0416	26x1 double	26x1 double	26x1 double	26x1 double	26x1 double	26x1 double	'-approxEnt'
10	0.0575	25x1 double	25x1 double	25x1 double	25x1 double	25x1 double	25x1 double	'+-approxEnt'

Figure 6.9: Output of wrapper model execution with information gathered over each of it's clustering iterations.

Figure 6.9 provides the results of a wrapper model's clustering iteration at each row. This information includes:

- Type of uterine event at column 2, where fetal movements are identified differently than other contractile events;
- Cluster assigned to each uterine event in third column;
- cluster centers in the forth column;
- Tracking of six clustering evaluation criteria considered on columns 5 to 10;

- Columns 11 to 16, keep track of six clustering evaluation criteria for each of the features being used;

- Column 12 shows features being added or removed.

Given that there is some lack of knowledge of what criterion may best define a good cluster in the domain considered, multiple criterion are saved at each iteration, although only one is used to determine the worse feature. In the example presented the order at which features were being removed was according to Je column criteria, which is the sum of squared errors. Of course, the criteria defined the order at which features were removed, but all are removed at some point. Furthermore they are added back by default if no negative effect is found in the accuracy of fetal movements classification, which makes the criteria pointless. However, as an academic work, studying these criteria evolution and correlating it with the clustering results found seems a crucial step for understanding what best evaluates extracted features.

At the end of the wrapper models execution the table presented in Figure 6.10 is created providing the clustering results of each uterine event.

12981x10 table

	1 type_contrac	2 id_contrac	3 id	4 id_bipSignal	5 patient	6 pCase	7 bipolar	8 weekGest	9 weekDel	10 clust
160	1	4801	22	219	10	1	9	29.7143	41.5714	4
161	1	4802	22	219	10	1	9	29.7143	41.5714	2
162	1	4803	22	219	10	1	9	29.7143	41.5714	3
163	1	4804	22	219	10	1	9	29.7143	41.5714	4
164	1	4805	22	219	10	1	9	29.7143	41.5714	1
165	1	4806	22	219	10	1	9	29.7143	41.5714	4
166	1	4807	22	219	10	1	9	29.7143	41.5714	1
167	1	4808	22	220	10	1	10	29.7143	41.5714	2
168	1	4809	22	220	10	1	10	29.7143	41.5714	2
169	1	4810	22	220	10	1	10	29.7143	41.5714	2
170	1	4812	22	220	10	1	10	29.7143	41.5714	2
171	1	4813	22	220	10	1	10	29.7143	41.5714	3
172	1	4814	22	220	10	1	10	29.7143	41.5714	2

Figure 6.10: Table outputted from Wrapper model execution showing final clustering result of each uterine event.

This last table doesn't hold much information and it's rather for analysis of the results obtained. Each row shows an uterine event, some profiling information and the cluster assigned to it.

The program explained across this section performs all the procedures to manipulate our data and collect results from it. Yet, inspection of how the signals are evolving with these procedures and of the results obtained is assessed by another tool developed detailed in the next section.

6.3 Visualization Tool

This project deals with a few problems for which the author found non-trivial the best solution's definition. It is not obvious that all preprocessing procedures applied have a positive outcome on improving uterine events characteristics. In the other hand, mere clustering statics over an unsupervised classification, for which the problem of validation has already been thoroughly discussed, may present difficulty for further work evolving the study of this solution. As such, it was implemented a visualization tool that allows inspecting the procedures output and a few of the results obtained by the methods used in this thesis.

The tool was implemented in Matlab and designed to allow comparing different methods, debugging the implemented procedures and inspecting the final results of this work. It, of course, does not process the data, as it can take many hours to days to run the full preprocessing, feature extraction and selection, and classification procedures to all the signals from Iceland database. That being the case, visualization tool's features unlock as output data files from the program detailed in 6.2 are generated. Meaning all main flow blocks of the latter stated program should be executed to allow using all features the visualization tool box provides.

The visualization tool can be divided in two parts, which in the program are referenced as tabs. The first, was implemented to perceive and adjust the initial preprocessing procedures prior to contraction detection, while the second contains contractions time and spectral representations and the clustering results. Figure 6.11 shows the tool's first tab.

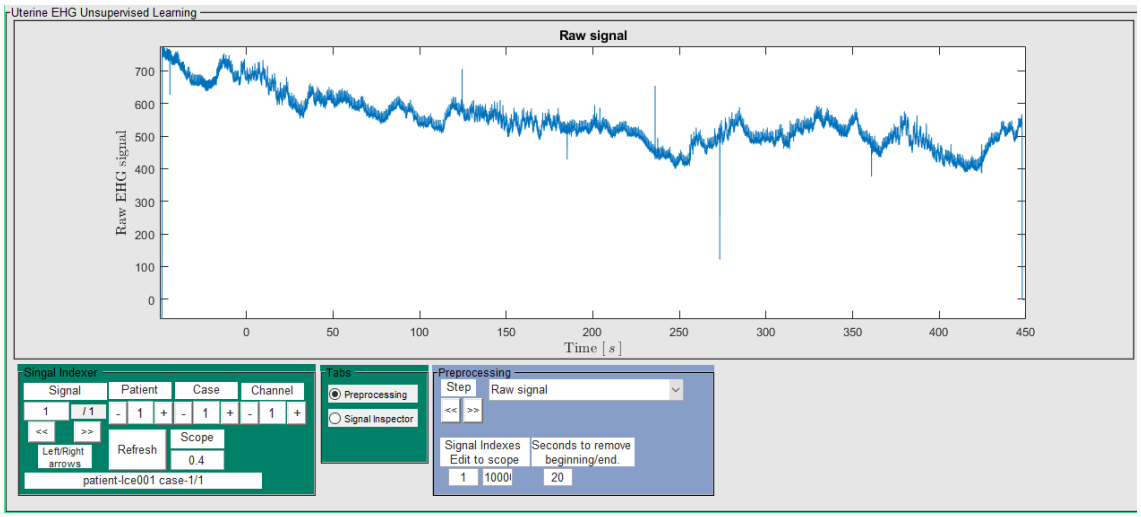


Figure 6.11: Preprocessing part of visualization tool.

The tool is composed of a plot window and the menus at the bottom, which contain all the tool's features. The first menu, *Signal Indexer*, allows changing the visualized signal either by patient, patient's visit (case), and bipolar channel. Scope edit box is not used in this first part. The second menu *Tabs*, is used to choose between *Preprocessing* and *Signal Inspector* parts of the visualization tool. The third menu allows controls the

preprocessing procedure to visualize (among those prior to contraction detection) and has a scope option, where the begging and ending samples can be specified. Figure 6.12 shows the preprocessing steps taken into account in this tab. The preprocessing tab is fairly simple, yet it was useful in assessing the results and adjusting parameters for the steps.

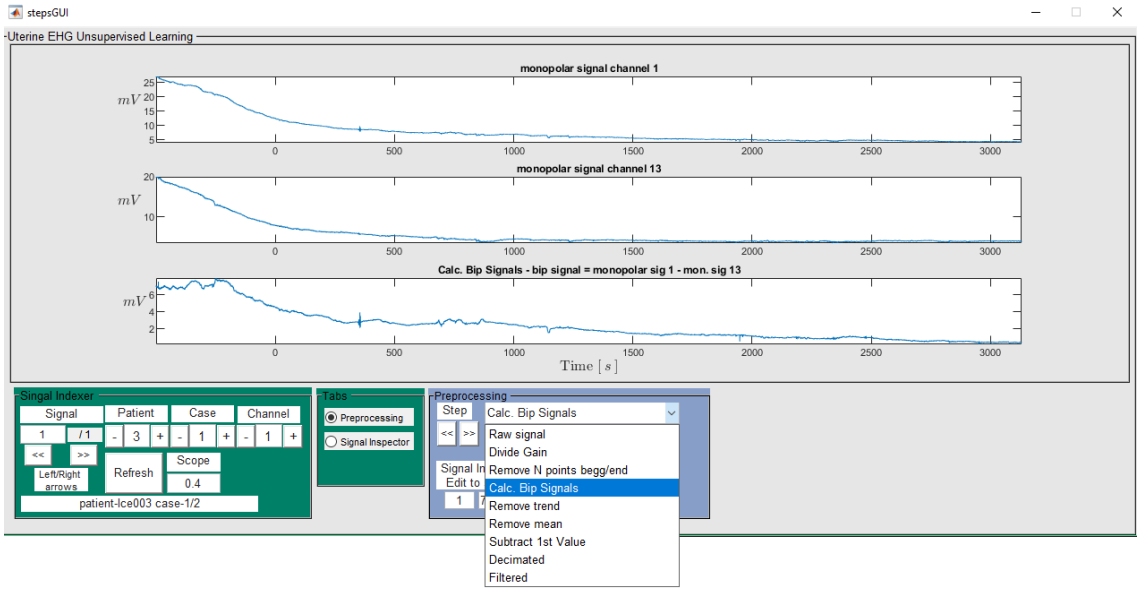


Figure 6.12: Preprocessing part of visualization tool. The drop-down list contains the preprocessing procedures that can be visualized, containing: 1-divide machine gain; 2-remove points from beginning and ending of signal, due to the amount of corrupted signals from moving electrodes when the test is starting and ending; 3-compute bipolar signals; 4-Remove signal tilt; 5-center signal to have zero average (this step was omitted from initial preprocessing description as it is not necessary); 6-subtract first value to the whole signal, for it to start at zero; 7-Result of Decimation; 8-Result of filtering to the band of interest.

The second tab is referred to as the *Signal Inspector*, which already works over the detected contractions instead of the whole signal. The bottom menus are as illustrated in Figure 6.13. The first two menus are kept from the preprocessing tab adding the difference that, now, it is scope edit box in *Signal Indexer* menu that provides features zooming in or out the contractions.

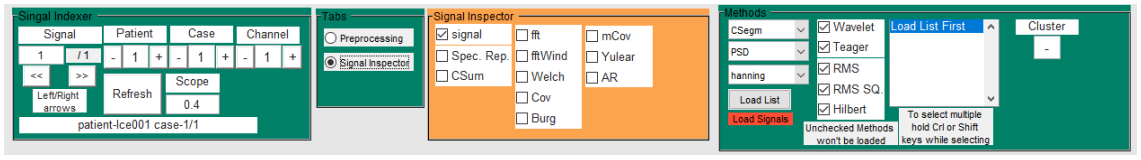


Figure 6.13: *Signal Inspector* Menus.

SignalInspector menu is used to choose whether it is visualized the contraction in the time domain (Figure 6.14), it's spectral representation (Figure 6.15) or the cumulative sum of the spectrum (Figure 6.16). The other check boxes refer to the spectrum estimators to use.

Methods menu contains combinations of the methods earlier discussed, namely, spectral representation (PSD, PS, LS or LSD), contraction and fetal movements detection methods and a window, which is used by Welch method to estimate spectral representations. After, choosing the methods to use, which determine what files the program is going to load when clicking '*Load List*' button.

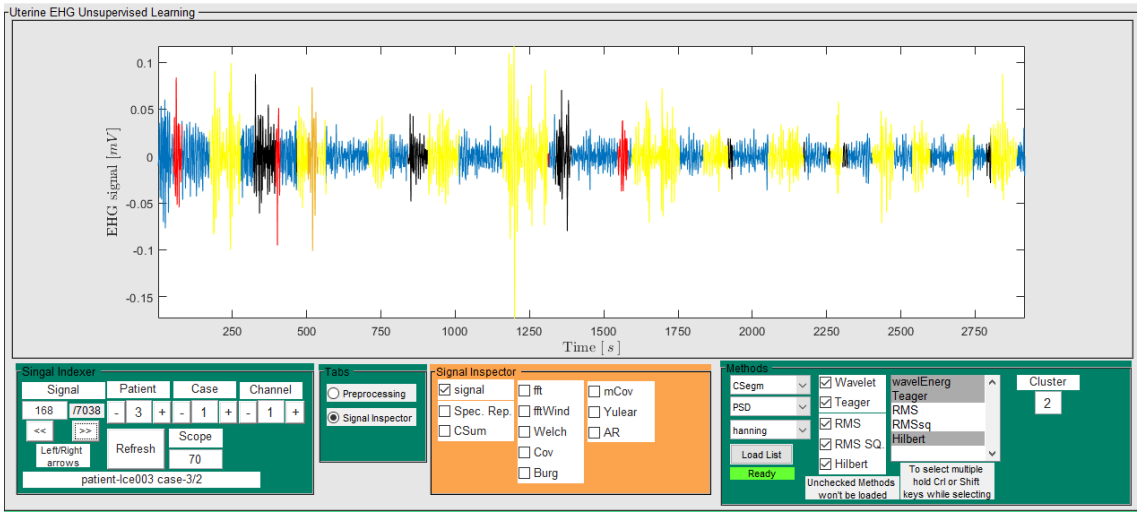


Figure 6.14: Visualization Tool - *Signal Inspector*: Here the scope was set to 70 times the signal length (in *Signal Indexer* menu). in *Methods* menu three contraction detection methods are chosen, namely, by wavelet energy (red segments), Teager energy operator (black segments) and Hilbert-Huang (yellow segments). As such in the defined scope all contraction detected by these methods will appear with different colors. The small orange segment is the selected one.

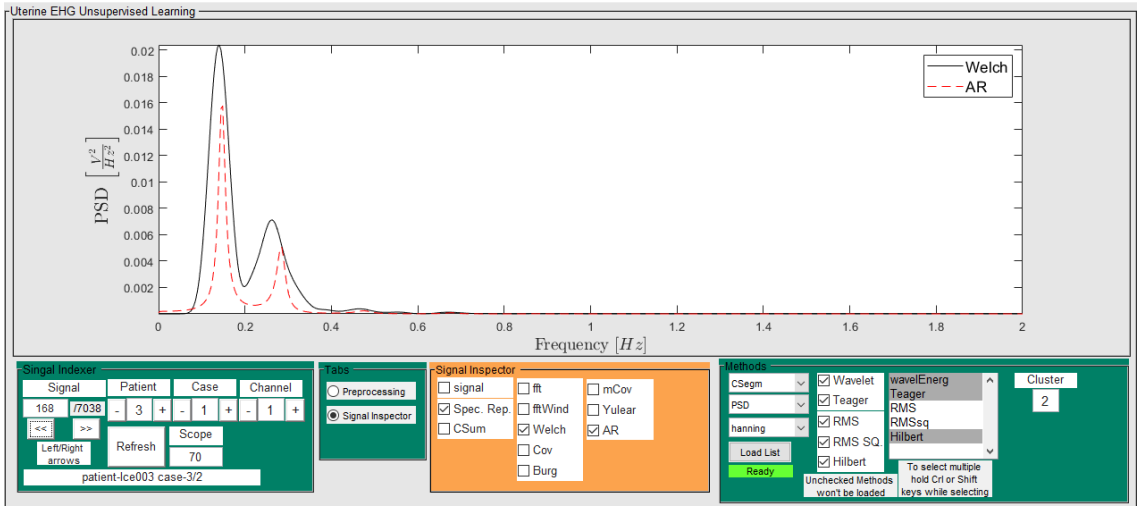


Figure 6.15: Spectral representation of the selected contraction segment. Here, only the PSD representations were loaded, and the figure shows it's estimation according to Welch and autoregressive (AR) estimators. Although many methods of contraction detection are selected, the contraction segments iterated are those with respect to the first oneto be selected (subsequent are selected by holding CTRL key).

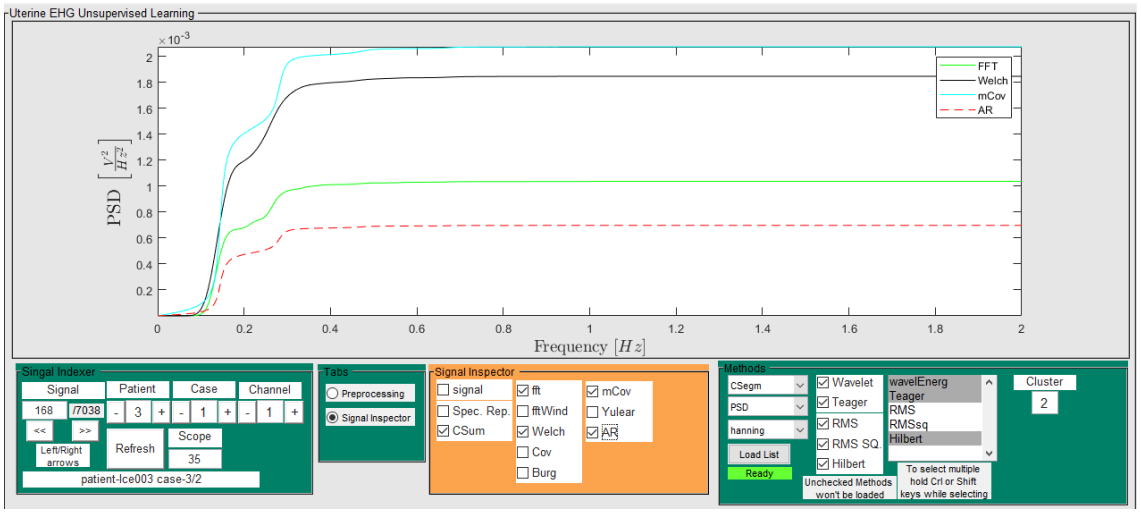


Figure 6.16: Cumulative Sum of the power representation. This plot simply integrates the estimated spectral representation to allow a visual inspection of how the energy evolves in the signal.

The tool allows the selection of more than one contraction detection methods, in order to compare the results of each method. The selected contraction is represented in orange and it's cluster is represented in the *Methods* menu according to the classification through the wrapper model described subsection 5.4.2.

Signal Inspector tab is meant to, not only inspect which methods seem to best represent the signal, but to allow visual validation of the clusters attained. A limit presented to the analysis of this signals is also the discrepancies among results from different contraction detection methods which, along clustering validation, presented the core motives for the implementation of this interface. As such, the tool was designed in a way that would allow an easy comparison between methods (exmaple in Figure 6.17) and would allow the inspection of the signal in both time and frequency domains when analyzing cluster results.

As illustrated in Figure 6.17, although the methods consider there is a contraction in the illustrated segment, there was a disagreement over the starting and stopping samples.

6.4 Results

This thesis contribution regards an unsupervised classification solution of extracted segments rated as uterine contractile events by Uterine Explorer [4] project's criteria.

As such, the results inspected in this section relate to feature selection and classification procedures. Processing procedures applied to the signals are not included here as there were no rigorous tests over of how well they performed. The processing procedures applied were decided upon results from other contributions to UEX project [4] and their study in this thesis concern producing a good protocol standard to obtain more satisfactory results.

Results will consider the execution of the program detailed in subsection 6.2 considering:

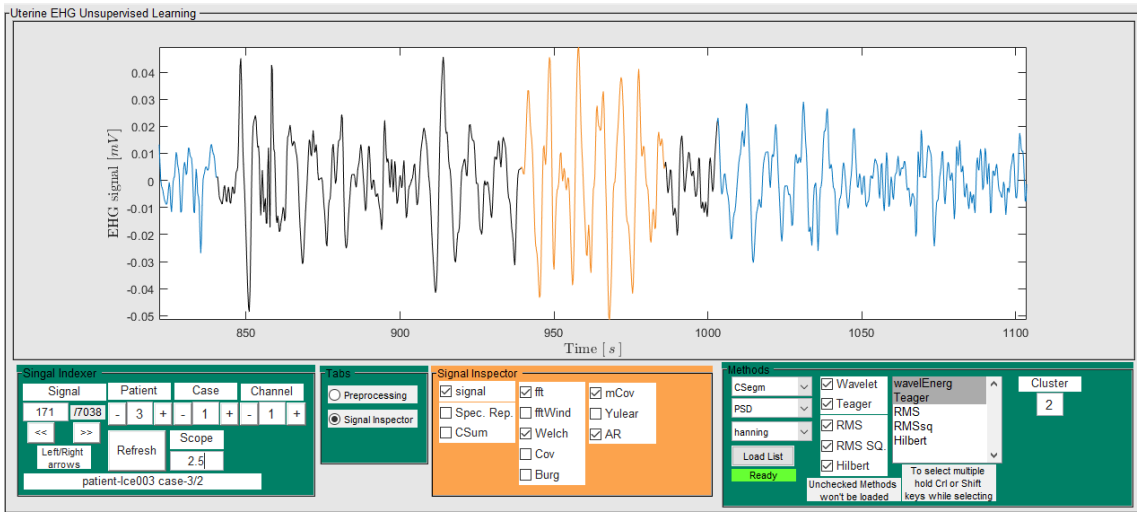


Figure 6.17: *SignalInspector* as a way to compare contraction detection methods. The orange signal segment a contraction detected using wavelet energy and the black segment represents a segment detected using Teager energy operator.

- Data downsampled to 4Hz;
- Signals components are filtered to be confined in a frequency band of 0.1 and 1Hz;
- Uterine events are extracted using wavelet energy method.
- Frequency domain representation of signal will be described by it's power spectral density (PSD) and estimated by welch using a Hanning window;

Feature sets used in the classification process include:

1. Uterine events represented in time domain interpolated to present a size of 512 samples
2. Power spectral density
3. A number of computed features earlier enumerated in subsection 6.2.1 represents the third set.

According to this description of considered data the results of PCA and wrapper model, along with their interpretations are provided in the following subsections.

6.4.1 Principal Component Analysis

To assess the capabilities of each of the sets earlier referred in this section (6.4), a PCA algorithm was applied to each and the number of used features was reduced according to a rate provided by their eigenvalues.

After applying PCA the eigenvalues were ordered ascendantly and only features above a given threshold were considered. The applied thresholds for the first and second sets

was 10%, while for the third it was 20%. These values were based on the amount of initial features of each set to avoid reducing that amount either too much or too little. Furthermore, visual inspection showed these values were accurate in that eigenvalues above those thresholds held exponential increase in their overall weight.

The described procedure allows choosing a lower dimensional transformation matrix which captures more relevant information according to the an eigenvalue analysis and simplifies our data. Subsequently a k-means clustering procedure was applied to the transformed and reduced sets. Figure 6.18 shows the distribution of contractions and fetal movements among clusters.

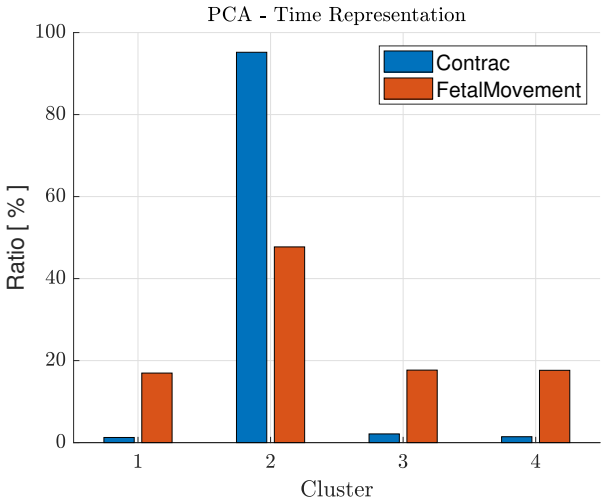


Figure 6.18: Distribution of the interpolated and transformed time-domain fetal movements and contractions among the clusters.

The result illustrated in Figure 6.18 is near opposite to the one intended as all the contractions were clustered together and the fetal movements were divided among the other clusters. What might have caused this result is the fact that fetal movements are normally smaller than other contraction, being more affected by the interpolation. Thus, their loss of characteristics may create a domain of signals harder to classify that work as outliers.

Without the interpolation, the same PCA and k-means procedures were applied to the other feature sets. The result of contractions and fetal movements clustering are represented in figure 6.20 and 6.20

Next feature set is the spectral representation of uterine events, which result is represented in Figure 6.20.

It is observable in Figure 6.20 that PSD feature set failed to differentiate fetal movements from other classes, but rather they seem similarly distributed among them. This result may favor the hypothesis that spectral features alone are not sufficient to discriminate all classes and do not seem to work on fetal movements. Furthermore, the principal component analysis of signals might not offer an accurate or at least proper way to address

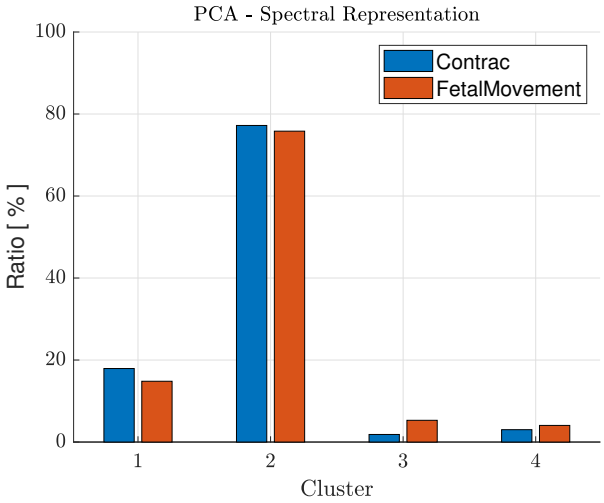


Figure 6.19: Distribution of the transformed PSD of fetal movements and contractions among the clusters.

our classification problem specifically.

Last feature set aggregates computed features and the classes distribution obtained is represented in figure 6.20.

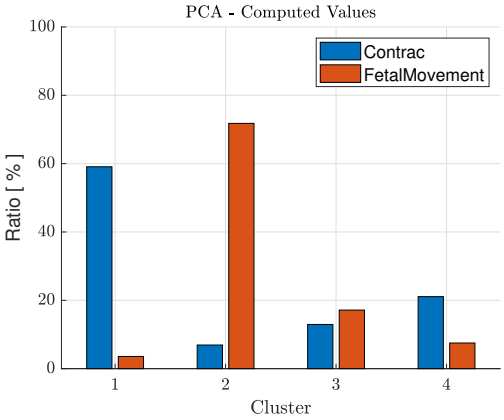


Figure 6.20: Distribution of the transformed computed features of fetal movements and contractions among the clusters.

The computed features set was able to distinguish fetal movements from other classes quite well. Fetal movement class has 71.78% correctly classified fetal movements and less than 7% misclassified contractions. This is an important result as it shows how PCA was not able to provide accurate solutions using neither time nor spectral characteristics separately. However, computed features holding both time and spectral information, along non-linear characteristics largely improved the result.

6.4.2 Wrapped Clustering Algorithm

The first test performed to the wrapped cluster algorithm uses synthetic data. The data was created by selecting three different contractions and one fetal movement. These were cut down to obtain signals equal in length and 150 versions of each were generated by a Gaussian distribution parameterized using their average and standard deviation. Subsequently spectral estimation and computation of features procedures were applied to the synthetic signals.

The computed features from the synthetic data were then provided to the wrapper model for which execution the fetal movement classification accuracy was optimized.

As described in subsection 5.4.2, the algorithm starts with all features and removes on by one re-adding those which are prejudicial to the fetal movements classification. After testing out all features, the algorithm tests re-adding the eliminated features. Adding and removing a feature is represented, in the results, plot by '+' and '−', respectively. The results are as shown in Figure 6.21.

This test was performed to attempt assess whether optimizing the classification of one class would benefit the others and offer some overall validation of the approach. The test showed a really good evolution of the clustering results in that the initially only 50 to 60% of the uterine event were correctly classified and optimizing the feature set reached around 100% accuracy for all classes. The positive results from this test are limited to it's context in that the signals were generated using a Gaussian distribution rather than a stochastic behavior.

The same logic is applied to the real data plots, with the difference that now only one class is known, so it is only plotted how well fetal movements are being classified (blue line) and the percentage of contractions that are being classified as fetal movements (red line). The algorithm attempts to lower the percentage of contractions being classified as fetal movements and to maximize fetal movement correct classification. Figure 6.22 shows the evolution in clustering results when wrapper model approach was applied to the computed values from real uterine events set.

The results presented in fig. 6.22 show accuracy for fetal movements classification is improved from 72.64% to 88.94% and there is also a decrease in other contractions being classified as fetal movements (FM) from 7.09% to 5.08%.

The models implementation, and the way it was parametrized for this execution in particular, assures current iteration is not more than 4% worse in correctly classifying FM nor 6% worse in wrongly classifying others as FM. When reaching those threshold the feature subset modification will be reversed. That being the case a positive outcome from the wrapper model is very probable and doesn't necessarily say much about whether the subset really improved it's capabilities.

The presented case in fig. 6.22 shows a good differentiation of fetal movements right from the first iteration, which shows the select initial set performed well. This characteristic creates a rather narrow optimizing band for the wrapper model's execution to work with,

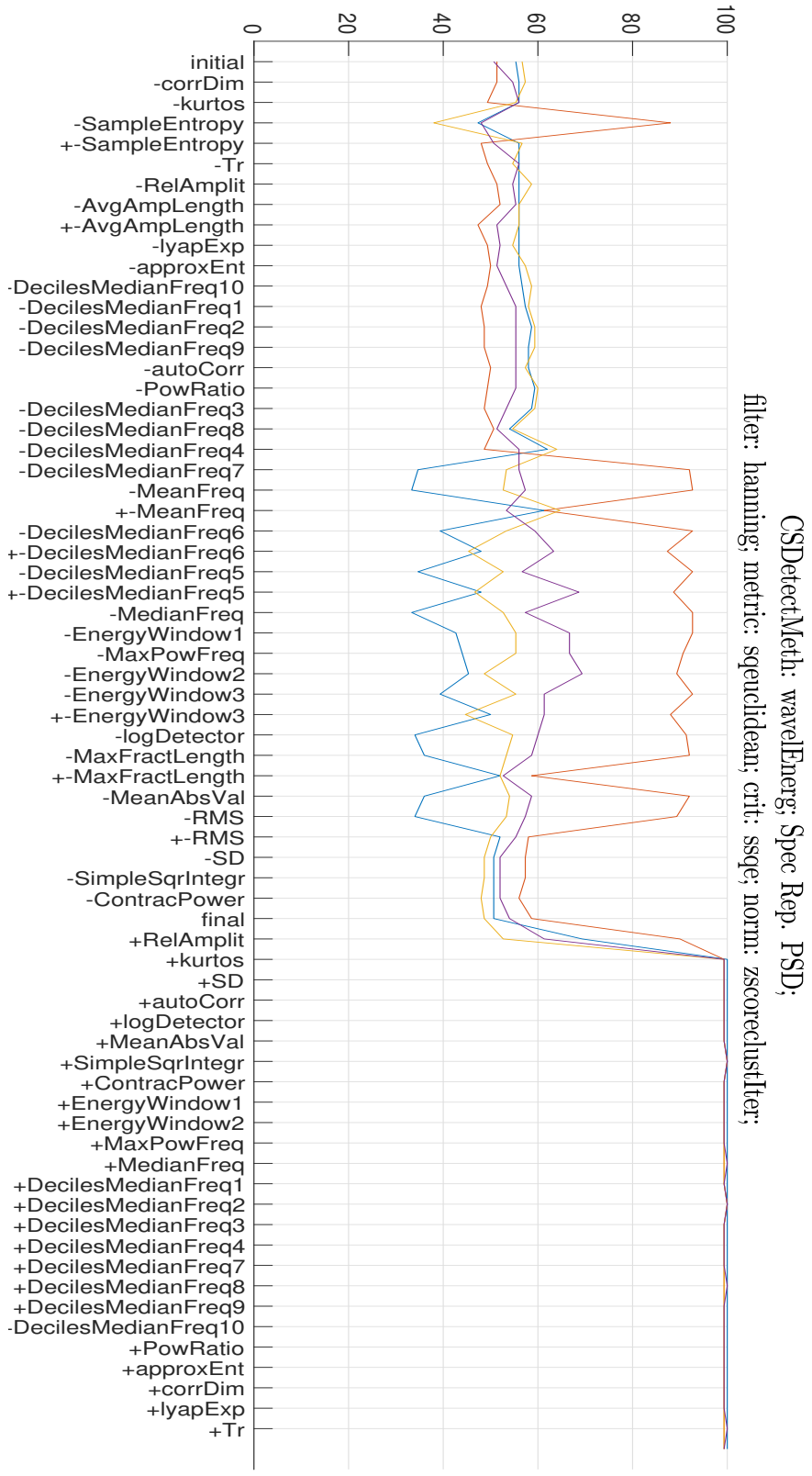


Figure 6.21: Wrapper model applied to synthetic data. Each line represents the percentage of correctly classified signals and the features being added and removed are being represented with a + or −, respectively.

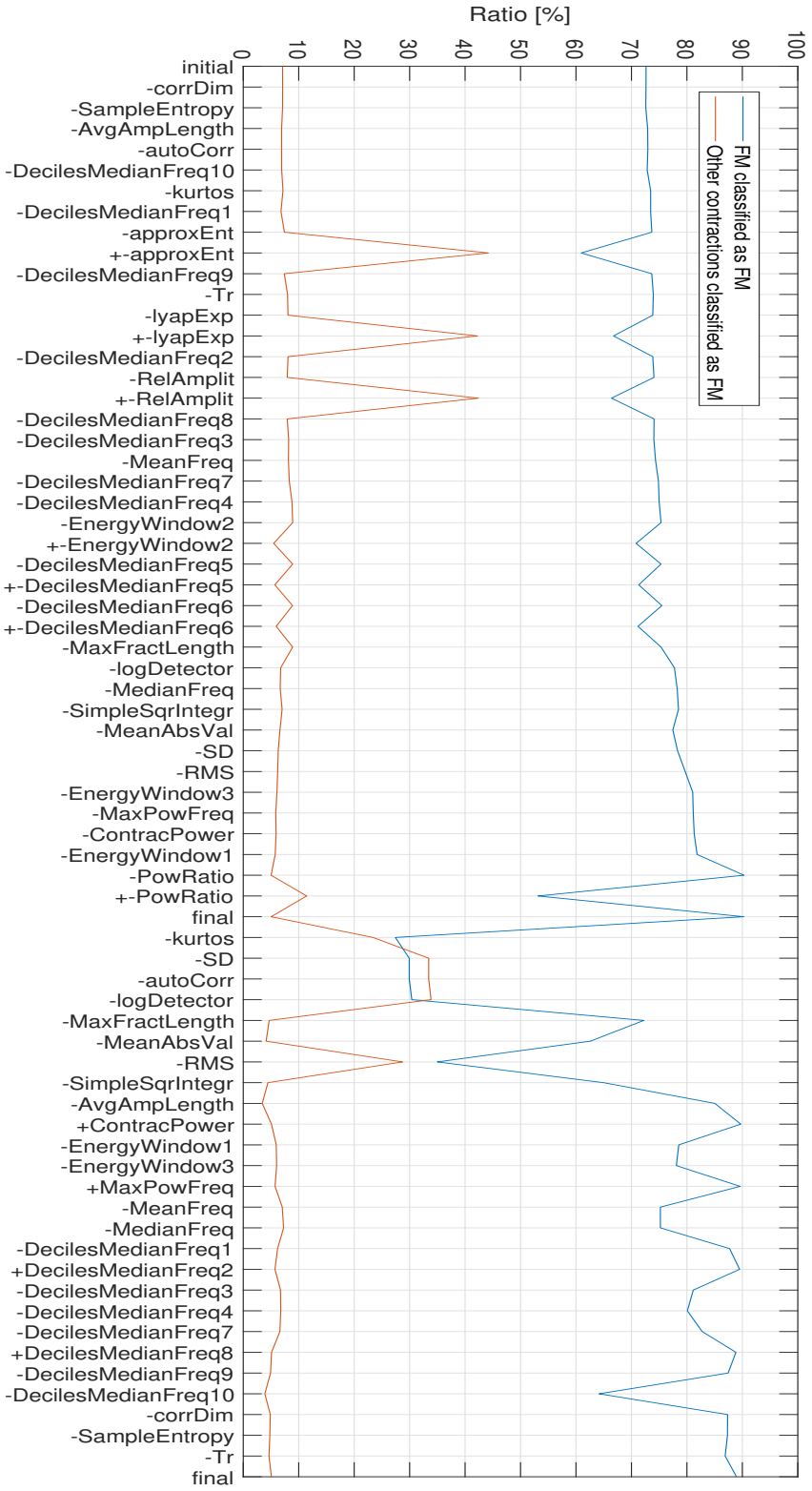


Figure 6.22: Iterations information of wrapper model. Blue line refers to the percentage of fetal movements being correctly classified, while the red line refers to the percentage of contractions being classified as fetal movements.

and less pronounced effects. However, the results obtained are quite satisfactory for a few reasons. The first refers to the the sequential backward selection (SBS) part of the execution's stability, implying subsets generated are gradually better in a constructive manner and not by chance. Second is the way re-added features affect the sequential forward selection (SFS) execution, disrupting the results. Naturally, when considering a higher number of features, the ominous ones influence is deluded. When the latter features are added in a smaller subset as the ones considered in SFS their effect is really visible and shows their removal was an appropriate choice.

On another note, the resulting subset from SBS actually defined the best result obtained of 89.51%, whereas the final one was only 88.94%, meaning SFS had a negative impact on optimizing the feature set. However, from other experiences SFS execution showed it's potential to reconsider highly valuable features were removed by less fortunate clustering results in SBS. This may occur because clustering centroids are initialized randomly. Thus, for high dimensional data without well defined class characteristics ultimately leads to local optima solutions rather than deterministic ones. To approach this problem, each iteration executes five clustering attempts and only keeps the best one according to sum-of-squared errors criteria in the given example. However, mistakes are possible and SFS, showed an overall positive contribution for the solution in across tests performed, at cost of computational resources and, possible reduction of accuracy made possible by the aforementioned 4% and 6% tolerance on approving negative impacting features.

Summing up the result is positive as a successful execution of the wrapper model should be defined a stable and increasing evolution of the accuracy in SBS and a low number of additions in SFS with high disruption of the clustering results by the re-added features.

6.4.3 Selected Subset Features

From wrapper models execution the optimized feature subset defined included: Lyapunov Exponent Approximate Entropy, RelAmplit, Second Energy Window, Power Ratio, Contraction Power, Deciles Median Freq 2,5,6 and 8, and Maximum Power Frequency. To test the subset of features, it was then applied PCA and k-means clustering procedure, which result is represented in Figure 6.23.

When compared to fig. 6.20, the result presented in fig. 6.23 shows an increase in correctly assigning fetal movements from 71.78% to 79.82% and a decrease of incorrectly assigning other uterine events as fetal movements from 6.92% to 5.10%. As such, the subset of features also improved the results for this procedure in the matter of improving fetal movements correct classification accuracy.

Another important aspect to analyze over the resulting cluster is how the obtained contractions are distributed over pregnancy stages. This inspection is an important test over the obtained clusters as it is closely related to the objective of identifying the different contractions in play. With the ability to label contractions the next step should be analyzing their evolution along pregnancy both in occurrences and how the classes are distributed

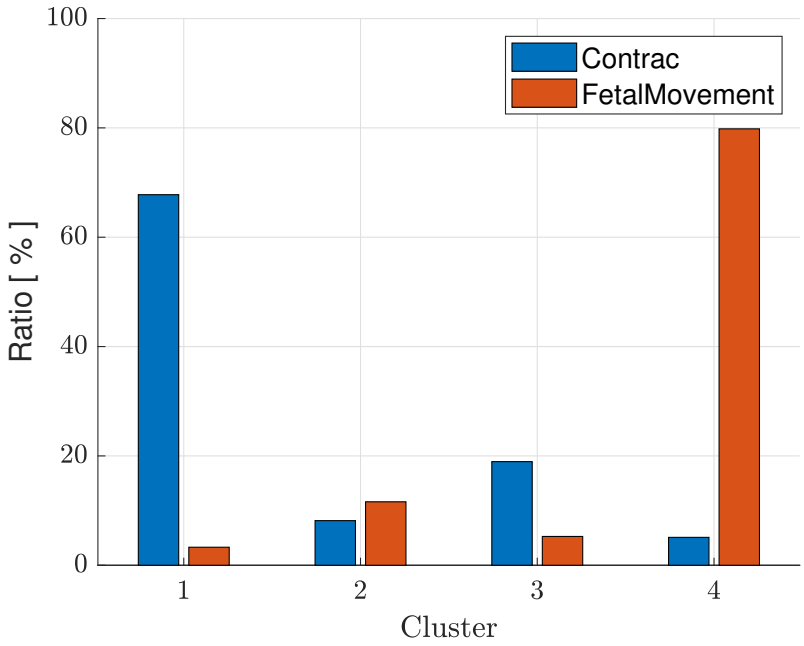


Figure 6.23: Distribution among clusters of the computed features selected by wrapper model and transformed by PCA algorithm.

along different stages of pregnancy. To perform this test, each contraction is associated to it's respective week of gestation and cluster obtained in the wrapper model's execution. The clusters distribution for each week is represented in Figure 6.24.

The result demonstrated in fig. 6.24 shows that the acquired clusters don't exhibit an obvious preference for contractions at any particular stage of pregnancy. The only cluster that may be better represented on later stages is cluster 3, yet it is also present at all stages and it's evolution is not as pronounced as it would require to take proper conclusions.

Results on the evolution of contractile activity is incomplete considering the test's gestational week alone, since it is not the same for all patients and it is more connected to how close delivery data is. As such, the same results is shown bellow (fig.

Looking at the result presented in fig.

As a last analysis of clusters distribution among different pregnancy stages and with respect to pregnancy stage a stacked graph in function of the stated weeks is represented in Figure 6.26.

The three-dimensional plot, although harder to interpret provides a more accurate representation of the distribution as it must be analyzed in function of both recording's respective gestational week and delivery week of it's patient case. This is justified by the evolution of contractile activity and efficiency of contractions nearing labor, described for Alvarez waves and Braxton Hicks in subsection 3.1.2. As such, connecting gestational week and delivery adds a context to the actual proximity to delivery.

Inspecting fig. 6.26, cluster 3 shows it's more representative presence over earlier term deliveries and less common on recordings that are close to labor. The event seems to skip a

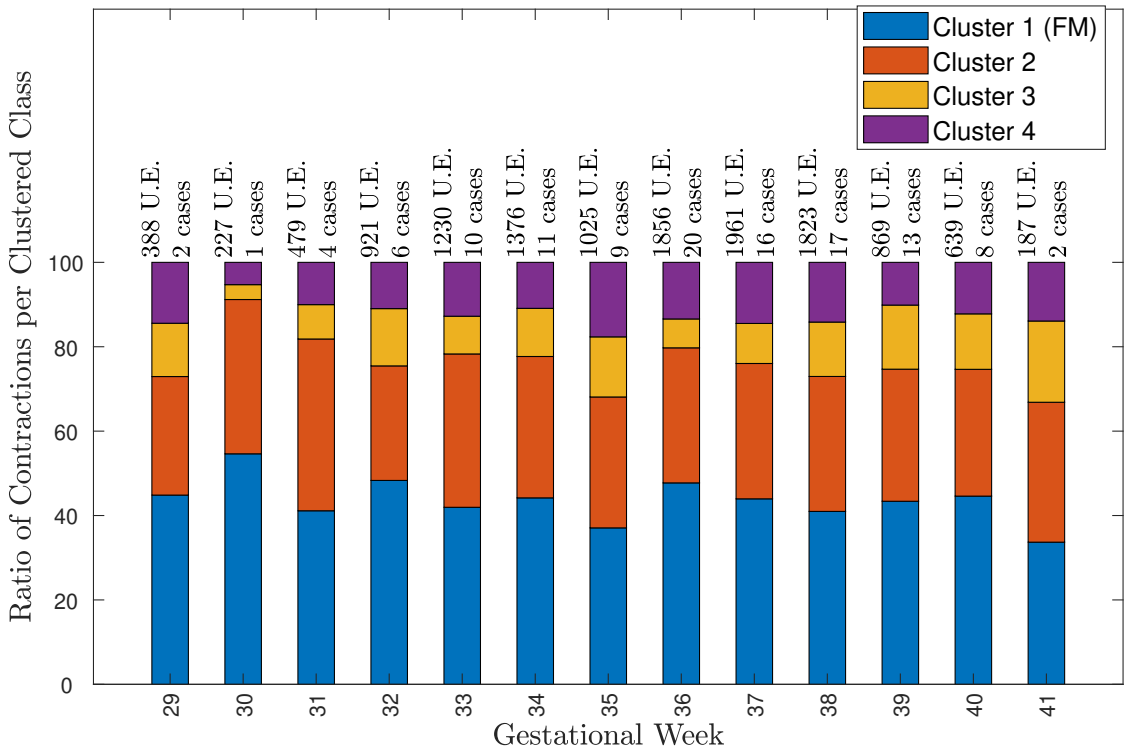


Figure 6.24: Stacked Graph representing percentage of each clustered class associated to each gestational week. The weeks are not rounded up, meaning a new week only ends at day one of posterior week. Text at the top of each bars indicates number of uterine events (U.E.) and patient cases considered for that week.

time period when approaching labor, being back again at labor recordings. Again, cluster 2 and 4 seem homogeneous along the tests, although cluster 1 seems to have slightly less weight on late recordings nearing and during labor.

In summation, a robust observation over the obtained clusters evolution along gestational and delivery weeks was not made possible by these results. Only cluster 3 seemed to hold discriminative capability for a pregnancy stage and results are inconclusive.

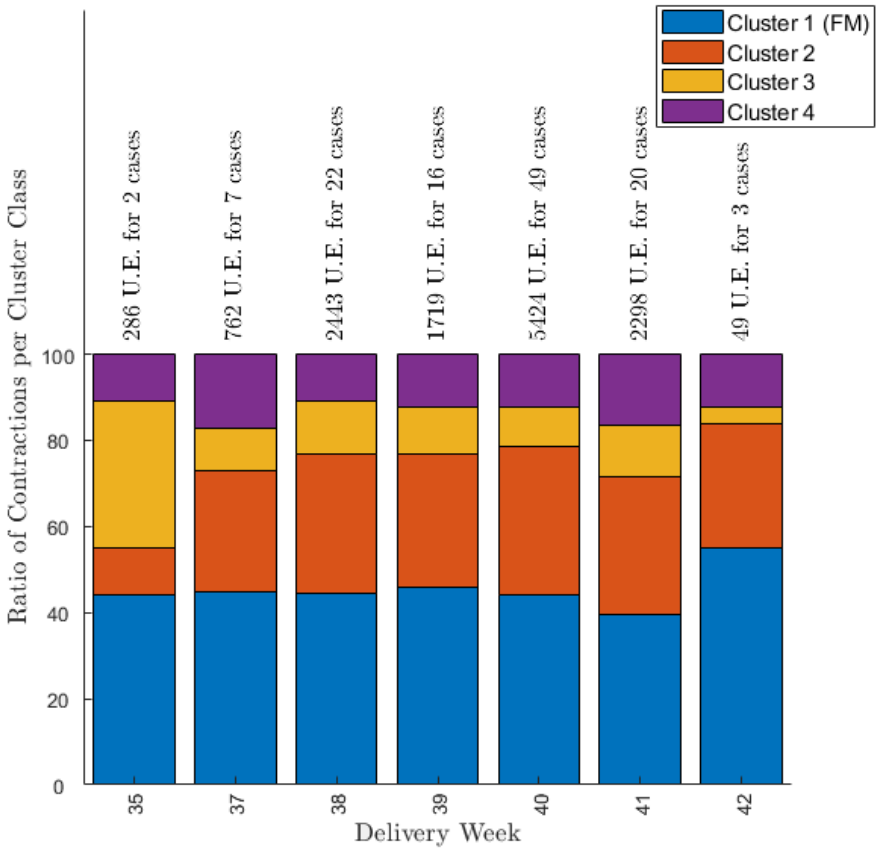


Figure 6.25: Stacked Graph representing percentage of each clustered class associated to each delivery week. The weeks are not rounded up, meaning a new week only ends at day one of posterior week. Text at the top of each bars indicates number of uterine events (U.E.) and patient cases considered for that week.

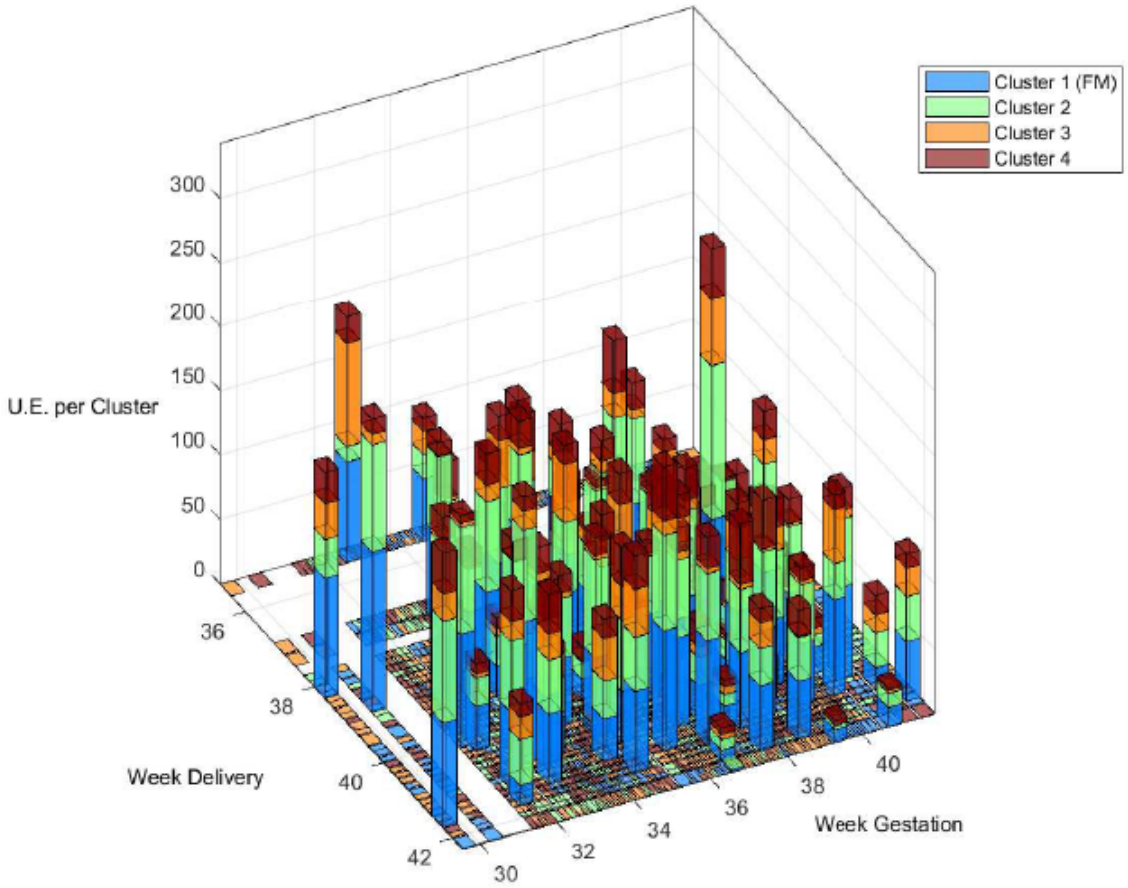


Figure 6.26: Stacked Graph representing number of contractions at each clustered class associated to each gestational week and delivery week.

CHAPTER 7

Conclusions

This thesis attempts to perform an unsupervised classification of uterine contractile events. It offers a contribution to a wider project the Uterine Explorer [4], which defines a number of processing procedures designed to improve the study of EHG signals, specifically. To such end, a study on these procedures parametrization was performed to allow standardizing a protocol that optimizes the input for the classification process.

The study of which EHG characteristics should be considered in the classification process was one of the main focuses of this work and one of it's biggest obstacles. To approach this problem, a number of features reported as useful from other works were considered and feature selection methods were employed to filter out ambiguous or less relevant features.

A characteristic of this work that allowed the approaches used was the fact fetal movements were extracted from in a separate procedure which allowed the assumption of knowing a class à priori. This characteristic allowed to infer from PCA and k-means procedure results, that neither time nor frequency characteristics alone held the capability of classifying or distinguish uterine events. Furthermore, the features chosen as the initial set held much better results alone without any feature selection method.

The focus of this work in feature selection method is justified by the characteristic of unsupervised classification procedures. Applying a statistical method such as a clustering procedure, will optimize a given criteria to best divide the signals into classes. However, the clustering procedure is contextualized by the features provided and even if a criteria is optimized to create very well concise and unambiguous clusters, it does not mean it is working towards the desired solution if the features simply don't influence it so.

There is not much knowledge about the uterine events to be identified, given that the signals are resultant of stochastic biological systems behavior. Furthermore, there are some studies attempting a similar goal, yet there is still not a concise solution to which features best characterize uterine events domain and distinguishes them. As such, this work approaches the problem by creating a wrapper model procedure using clustering algorithms optimized by accuracy in the classification of fetal movements in order to identify features that best define this class behavior. As such, the features subset is optimized according to the supervised characteristic of the problem specifically, yet the classification procedure

itself at each iteration of the wrapper model is unsupervised and unbiased. Optimization of the subset of features towards the described goal was verified with good results.

A test on synthetic data generated by a Gaussian distribution with characteristics from four different contractions simulated the wrapper model execution in a situation where again only one class is optimized, yet it was verified that all classes reached roughly 100% accuracy. This result doesn't prove the model is correct, since signals were defined by a Gaussian distribution, which is not transposable to the stochastic behavior verified in the real contractile events, but the clustering procedure was optimized by the models execution.

When applying PCA and k-means clustering the the new subset of features the results were also improved in the matter of accurately distinguishing fetal movements with the subset provided by the wrapper model. Thus, the optimization was successful, although under the premise that accurately defining one of the classes will generate a feature subset that better defines the behavior of these signals, which may lead to discarding necessary features to distinguish other classes.

Inferences from the analysis of resulting classes occurrence evolution in function of pregnancy stage and week of delivery was not successful, as their evolution didn't show clear and statistically relevant results. The result shows Braxton Hicks might not be represented by an obtained class given that they only obtain a regular rhythm in the last two weeks of pregnancy as described in subsection 3.1.2.

The major difficulties found in the course of this work were produced by computation requirements and results validation given the lack of à priori class information knowledge.

Overall, the presented solution optimized feature set to increase classification of the known class up to a very good accuracy. It still maintained an unsupervised unbiased approach at the classification procedure.

7.1 Future Work

The unsupervised classification solution proposed in this thesis optimized a set of features, computed from extracted uterine events and their spectral representations, to which a clustering method was applied. Furthermore a visualization tool was implemented to allow analyzing each preprocessing step's effect, the obtained signals in time and frequency domains and the cluster to which each signal was assigned to. Regarding future work in this project, specifically, a major step towards the study of this signals would include a manual classification performed by a specialist and validation of cluster results, which was one of purposes a visualization tool was implemented. Creating a known database would allow better defining the uterine events extraction method, studying the characteristics of these signals that should be explored and validation of this work's and future solutions.

This thesis put more focus in perceiving which features would work better in the classification process, yet the clustering method applied should also have room for improvement.

Specifically, if using a clustering method, combination of metric, criteria and feature normalization method, proved to be a rather more complex and domain dependent problem than initially expected, and one for which this work could not provided a grounded proposal. To approach this problem, the distribution of data must be analyzed and a knowledge of the patterns that should be detected is necessary. In other words, agreeing upon a set of features it should be performed a study of how these are distributed along each class, since even if they are clearly distinct between different classes the used metric e.g. euclidean distance, might not capture how these should be distributed.

From the point a robust classifier is defined and is capable of generalization, patterns on pregnancy stages and difficulties should become the next field to explore.

On a final note, it should be attempted a standardization of the way EHG signals are collected. The lack of data creates low statistical relevance of the results found and becomes a bigger problem when analyzing specific cases such as preterm labor. A possibly viable solution might be found in data balancing methods such as Adaptive synthetic sampling approach (ADASYN) [29] and Synthetic minority over-sampling technique (SMOTE) [14] to overcome the problem of misrepresented classes.

Bibliography

[1] Cohen. "Clinical Assessement of Uterine Contractions." In: ().

[2] M. O. Diab, A. El-Merhie, N. El-Halabi, and L. Khoder. "Classification of uterine EMG signals using supervised classification method." In: *Journal of Biomedical Science and Engineering* 03.09 (2010), pp. 837–842. ISSN: 1937-6871. DOI: 10.4236/jbise.2010.39113. URL: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jbise.2010.39113.

[3] A. Alexandersson, T. Steingrimsdottir, J. Terrien, C. Marque, and B. Karlsson. "The Icelandic 16-electrode electrohysterogram database." In: *Scientific Data* 2 (2015), p. 150017. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.17. URL: http://www.nature.com/articles/sdata201517.

[4] A. Batista, S. Najdi, D. Godinho, C. Martins, F. Serrano, M. Ortigueira, and R. Rato. "A multichannel time–frequency and multi-wavelet toolbox for uterine electromyography processing and visualisation." In: *Computers in Biology and Medicine* 76 (2016). ISSN: 18790534. DOI: 10.1016/j.compbiomed.2016.07.003.

[5] J. Alberola-Rubio, G. Prats-Boluda, Y. Ye-Lin, J. Valero, A. Perales, and J. Garcia-Casado. "Comparison of non-invasive electrohysterographic recording techniques for monitoring uterine dynamics." In: *Medical Engineering and Physics* 35.12 (2013), pp. 1736–1743. ISSN: 13504533. DOI: 10.1016/j.medengphy.2013.07.008. URL: http://dx.doi.org/10.1016/j.medengphy.2013.07.008.

[6] L. Dill, R. M. Maiden, and R. Engineer. "The electrical potentials of the human uterus in labor." In: *American Journal of Obstetrics and Gynecology* 52.5 (2015), pp. 735–745. ISSN: 00029378. DOI: 10.1016/0002-9378(46)90180-9. URL: http://dx.doi.org/10.1016/0002-9378(46)90180-9.

[7] C. Marque, J. M. Duchene, S. Leclercq, G. S. Panczer, and J. Chaumont. "Uterine EHG Processing for Obstetrical Monitoring." In: *IEEE Transactions on Biomedical Engineering* BME-33.12 (1986), pp. 1182–1187. ISSN: 15582531. DOI: 10.1109/TBME.1986.325698.

[8] H. ALVAREZ and R. CALDEYRO. *Gynecology and Obstetrics*. Vol. 91. 1950. ISBN: 192962204X.

[9] W. F. Roberts, K. G. Perry, R. W. Naef, J. F. Washburne, and J. C. Morrison. "The irritable uterus: A risk factor for preterm birth?" In: *American Journal of Obstetrics and Gynecology* 172.1 PART 1 (1995), pp. 138–142. ISSN: 00029378. DOI: 10.1016/0002-9378(95)90102-7.

[10] I. Blickstein. "MULTIPLE Pregnancy Epidemiology, Gestation & Perinatal outcome." In: ().

[11] M Khalil and J. Duchene. "Uterine EMG analysis: a dynamic approach for change detection and classification." In: *IEEE Transactions on Biomedical Engineering* 47.6 (2000), pp. 748–756. ISSN: 00189294. DOI: 10.1109/10.844224. URL: http://www.ncbi.nlm.nih.gov/pubmed/10833849http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=844224.

[12] M. Chendeb, M. Khalil, D. Hewson, and J. Duchên. "Classification of non stationary signals using multiscale decomposition." In: *Journal of Biomedical Science and Engineering* 03.02 (2010), pp. 193–199. ISSN: 1937-6871. DOI: 10.4236/jbise.2010.32025.

[13] K. Horoba, A. Matonia, J. Jezewski, T. Kupka, and A. Gacek. "Analysis of Uterine Contraction Activity Using two Ways of Signal Acquisition." In: *XI Conference "Medical Informatics & Technologies"* April 2016 (2006), pp. 199–204.

[14] P. Fergus, I. Idowu, A. Hussain, and C. Dobbins. "Advanced artificial neural network classification for detecting preterm births using EHG records." In: *Neurocomputing* 188 (2016), pp. 42–49. ISSN: 18728286. DOI: 10.1016/j.neucom.2015.01.107.

[15] D. Alamedine, M. Khalil, and C. Marque. "Comparison of Different EHG Feature Selection Methods for the Detection of Preterm Labor." In: *Computational and Mathematical Methods in Medicine* 2013 (2013), pp. 1–9. ISSN: 1748-670X. DOI: 10.1155/2013/485684. URL: http://www.hindawi.com/journals/cmmm/2013/485684/abs/{\%}5Cnhttp://downloads.hindawi.com/journals/cmmm/2013/485684.pdf{\%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3884970{\&}tool=pmcentrez{\&}rendertype=abstract{\%}5Cnhttp://www.hindawi.com/journals/.

[16] C. K. Marque, J. Terrien, S. Rihana, and G. Germain. "Preterm labour detection by use of a biophysical marker: the uterine electrical activity." In: *BMC Pregnancy and Childbirth* 7.Suppl 1 (2007), S5. DOI: 10.1186/1471-2393-7-S1-S5. URL: http://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/1471-2393-7-S1-S5.

[17] C. Rabotti, M. Mischi, J. O. Van Laar, P. Aelen, S. G. Oei, and J. W. Bergmans. "Relationship between electrohysterogram and internal uterine pressure: A preliminary study." In: *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings* 2017.1 (2006), pp. 1661–1664. ISSN: 05891019. DOI: 10.1109/IEMBS.2006.259928. URL: https://www.hindawi.com/journals/cmmm/2017/7949507/.

[18] M. O. Diab, B. Moslem, M. Khalil, and C. Marque. "Classification of uterine EMG signals by using normalized wavelet packet energy." In: *Proceedings of the Mediterranean Electrotechnical Conference - MELECON* (2012), pp. 335–338. ISSN: 2158-8473. DOI: 10.1109/MELCON.2012.6196443.

[19] A. Diab. "Study of The Nonlinear Properties And Propagation Characteristics Of The Uterine Electrical Activity During Pregnancy And Labor." In: (2017). URL: https://hal.archives-ouvertes.fr/tel-01430773.

[20] T. Steingrimsdottir, V. Gudmundsson, C. Marque, J. Terrien, and B. Karlsson. "Abdominal EHG on a 4 by 4 grid: mapping and presenting the propagation of uterine contractions." In: *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007* (2007), pp. 139–143. DOI: 10.1007/978-3-540-73044-6_35.

[21] M. Hassan, J. Terrien, C. Muszynski, A. Alexandersson, C. Marque, and B. Karlsson. "Better pregnancy monitoring using nonlinear correlation analysis of external uterine electromyography." In: *IEEE Transactions on Biomedical Engineering* 60.4 (2013), pp. 1160–1166. ISSN: 00189294. DOI: 10.1109/TBME.2012.2229279.

[22] W. L. Maner, L. B. MacKay, G. R. Saade, and R. E. Garfield. "Characterization of abdominally acquired uterine electrical signals in humans, using a non-linear analytic method." In: *Medical and Biological Engineering and Computing* 44.1-2 (2006), pp. 117–123. ISSN: 01400118. DOI: 10.1007/s11517-005-0011-3.

[23] M. Hassan, J. Terrien, C. Marque, and B. Karlsson. "Comparison between approximate entropy, correntropy and time reversibility: Application to uterine electromyogram signals." In: *Medical Engineering and Physics* 33.8 (2011), pp. 980–986. ISSN: 13504533. DOI: 10.1016/j.medengphy.2011.03.010. URL: http://dx.doi.org/10.1016/j.medengphy.2011.03.010.

[24] G. Fele-Žorž, G. Kavšek, Ž. Novak-Antolič, and F. Jager. "A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups." In: *Medical and Biological Engineering and Computing* 46.9 (2008), pp. 911–922. ISSN: 01400118. DOI: 10.1007/s11517-008-0350-y.

[25] B. Moslem, M. Diab, M. Khalil, and C. Marque. "Combining data fusion with multiresolution analysis for improving the classification accuracy of uterine EMG signals." In: *EURASIP Journal on Advances in Signal Processing* 2012.1 (2012), p. 167. ISSN: 1687-6180. DOI: 10.1186/1687-6180-2012-167. URL: https://asp-eurasipjournals.springeropen.com/articles/10.1186/1687-6180-2012-167.

[26] J. Meng, L. M. Meriño, N. B. Shamlo, S. Makeig, K. Robbins, and Y. Huang. "Characterization and Robust Classification of EEG Signal from Image RSVP Events with Independent Time-Frequency Features." In: *PLoS ONE* 7.9 (2012). ISSN: 19326203. DOI: 10.1371/journal.pone.0044464.

[27] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau. "EMG feature evaluation for improving myoelectric pattern recognition robustness." In: *Expert Systems with Applications* 40.12 (2013), pp. 4832–4840. ISSN: 09574174. DOI: 10.1016/j.eswa.2013.02.023. URL: http://dx.doi.org/10.1016/j.eswa.2013.02.023.

[28] M. O. Diab, C. Marque, and M. A. Khalil. "Classification for uterine EMG signals: Comparison between AR model and statistical classification method." In: *International Journal of Computational Cognition* 5.June 2014 (2007).

[29] U. R. Acharya, V. K. Sudarshan, S. Q. Rong, Z. Tan, C. M. Lim, J. E. Koh, S. Nayak, and S. V. Bhandary. "Automated detection of premature delivery using empirical mode and wavelet packet decomposition techniques with uterine electromyogram signals." In: *Computers in Biology and Medicine* 85.December 2016 (2017), pp. 33–42. ISSN: 18790534. DOI: 10.1016/j.compbiomed.2017.04.013. URL: http://dx.doi.org/10.1016/j.compbiomed.2017.04.013.

[30] K. Horoba, J. Jezewski, A. Matonia, J. Wrobel, R. Czabanski, and M. Jezewski. "Early predicting a risk of preterm labour by analysis of antepartum electrohysterograhic signals." In: *Biocybernetics and Biomedical Engineering* 36.4 (2016), pp. 574–583. ISSN: 02085216. DOI: 10.1016/j.bbe.2016.06.004. URL: http://dx.doi.org/10.1016/j.bbe.2016.06.004.

[31] A. Lemancewicz, M. Borowska, P. Kuć, E. Jasińska, P. Laudański, T. Laudański, and E. Oczeretko. "Early diagnosis of threatened premature labor by electrohysterographic recordings - The use of digital signal processing." In: *Biocybernetics and Biomedical Engineering* 36.1 (2016), pp. 302–307. ISSN: 02085216. DOI: 10.1016/j.bbe.2015.11.005.

[32] Cohen, Barrie Hayes-Gill, Sarmina Hassan, Fadi G. Mirza, Sophia Ommani, Barrie Hayes-Gill, Molham Solomon, Raymond Brown, Barry S. Schifrin, and Cohen. "Accuracy and Reliability of Uterine Contraction Identification Using Abdominal Surface Electrodes." In: *Clinical Medicine Insights: Women's Health* (2012), p. 65. ISSN: 1179-562X. DOI: 10.4137/CMWH.S10444. URL: http://la-press.com/accuracy-and-reliability-of-uterine-contraction-identification-using-a-article-a3445.

[33] Z. Liu, D. Hao, L. Zhang, J. Liu, X. Zhou, L. Yang, Y. Yang, X. Li, and D. Zheng. "Comparison of electrohysterogram characteristics during uterine contraction and non - contraction during labor *." In: (2017), pp. 2924–2927.

[34] J. Alberola-Rubio, J. Garcia-Casado, G. Prats-Boluda, Y. Ye-Lin, D. Desantes, J. Valero, and A. Perales. "Prediction of labor onset type: Spontaneous vs induced; role of electrohysterography?" In: *Computer Methods and Programs in Biomedicine* 144 (2017), pp. 127–133. ISSN: 18727565. DOI: 10.1016/j.cmpb.2017.03.018. URL: http://dx.doi.org/10.1016/j.cmpb.2017.03.018.

[35] M. Hassan, J. Terrien, B. Karlsson, and C. Marque. "Interactions between uterine EMG at different sites investigated using wavelet analysis: Comparison of pregnancy and labor contractions." In: *Eurasip Journal on Advances in Signal Processing* 2010 (2010). ISSN: 16876172. DOI: 10.1155/2010/918012.

[36] M. O. Diab, C. Marque, and M. Khalil. "An unsupervised classification method of uterine electromyography signals: Classification for detection of preterm deliveries." In: *Journal of Obstetrics and Gynaecology Research* 35.1 (2009), pp. 9–19. ISSN: 13418076. DOI: 10.1111/j.1447-0756.2008.00981.x.

[37] R. Merletti and P. A. Parker. *Electromyography*. ISBN: 3175723993.

[38] C. Rabotti. *Characterization of uterine activity by electrohysterography*. april. 2010, p. 145. ISBN: 9789038622057. DOI: 978-90-386-2205-7.

[39] Y. Ye-Lin, J. Garcia-Casado, G. Prats-Boluda, J. Alberola-Rubio, and A. Perales. "Automatic identification of motion artifacts in EHG recording for robust analysis of uterine contractions." In: *Computational and Mathematical Methods in Medicine* 2014 (2014). ISSN: 1748670X. DOI: 10.1155/2014/470786.

[40] W. L. Maner and R. E. Garfield. "Identification of human term and preterm labor using artificial neural networks on uterine electromyography data." In: *Annals of Biomedical Engineering* 35.3 (2007), pp. 465–473. ISSN: 00906964. DOI: 10.1007/s10439-006-9248-8. URL: http://link.springer.com/10.1007/s10439-006-9248-8http://www.ncbi.nlm.nih.gov/pubmed/17226089.

[41] T. Y. Euliano, M. T. Nguyen, S. Darmanjian, S. P. McGorray, N. Euliano, A. Onkala, and A. R. Gregg. "Monitoring uterine activity during labor: A comparison of 3 methods." In: *American Journal of Obstetrics and Gynecology* 208.1 (2013), 66.e1–66.e6. ISSN: 00029378. DOI: 10.1016/j.ajog.2012.10.873. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.ajog.2012.10.873.

[42] P. Fergus, P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, and S. Iram. "Prediction of Preterm Deliveries from EHG Signals Using Machine Learning." In: *PLoS ONE* 8.10 (2013). ISSN: 19326203. DOI: 10.1371/journal.pone.0077154.

[43] S. Sim, H. Ryou, H. Kim, J. Han, and K. Park. "The 15th International Conference on Biomedical Engineering." In: 43 (2014), pp. 675–676. DOI: 10.1007/978-3-319-02913-9. URL: http://link.springer.com/10.1007/978-3-319-02913-9.

[44] J. Jezewski, K. Horoba, A. Matonia, and J. Wrobel. "Quantitative analysis of contraction patterns in electrical activity signal of pregnant uterus as an alternative to mechanical approach." In: *Physiological Measurement* 26.5 (2005), pp. 753–767. ISSN: 09673334. DOI: 10.1088/0967-3334/26/5/014.

[45] M. Lucovnik, W. L. Maner, L. R. Chambliss, R. Blumrick, J. Balducci, Z. Novak-Antolic, and R. E. Garfield. "Noninvasive uterine electromyography for prediction of preterm delivery." In: *American Journal of Obstetrics and Gynecology* 204.3 (2011), 228.e1–228.e10. ISSN: 00029378. DOI: 10.1016/j.ajog.2010.09.024. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.ajog.2010.09.024.

[46] B. Vasak, E. M. Graatsma, E. Hekman-Drost, M. J. Eijkemans, J. H. Schagen Van Leeuwen, G. H. Visser, and B. C. Jacod. "Uterine electromyography for identification of first-stage labor arrest in term nulliparous women with spontaneous onset of labor." In: *American Journal of Obstetrics and Gynecology* 209.3 (2013), 232.e1–232.e8. ISSN: 00029378. DOI: 10.1016/j.ajog.2013.05.056. URL: http://dx.doi.org/10.1016/j.ajog.2013.05.056.

[47] T. Y. Euliano, D. Marossero, M. T. Nguyen, N. R. Euliano, J. Principe, and R. K. Edwards. "Spatiotemporal electrohysterography patterns in normal and arrested labor." In: *American Journal of Obstetrics and Gynecology* 200.1 (2009), 54.e1–54.e7. ISSN: 00029378. DOI: 10.1016/j.ajog.2008.09.008. URL: http://dx.doi.org/10.1016/j.ajog.2008.09.008.

[48] C. Rabotti, M. Mischi, J. O. Van Laar, P. Aelen, S. G. Oei, and J. W. Bergmans. "Relationship between electrohysterogram and internal uterine pressure: A preliminary study." In: *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings* (2006), pp. 1661–1664. ISSN: 05891019. DOI: 10.1109/IEMBS.2006.259928.

[49] M. Diab. "Classification des signaux EMG utérins afin de détecter les accouchements prématurés." In: December 2007 (2007). URL: http://tel.archives-ouvertes.fr/tel-00410409.

[50] K. Kiasaleh. *Biological Signals Classification and Analysis (Lecture Notes in Bioengineering)*. 2015. ISBN: 9783642548789. DOI: 10.1007/978-3-642-54879-6. URL: http://www.amazon.com/Biomedical-Classification-Analysis-Lecture-Bioengineering/dp/3642548784/{\%}5Cnfile:///C:/Users/nogawa/SkyDrive/Mendeley/Kiasaleh-2015-BiologicalSignalsClassificationandAnalysis(LectureNotesinBioe.pdf.

[51] B. Moslem, M. Khalil, C. Marque, and M. O. Diab. "Energy distribution analysis of uterine electromyography signals." In: *Journal of Medical and Biological Engineering* 30.6 (2010), pp. 361–366. ISSN: 16090985. DOI: 10.5405/jmbe.768.

[52] R. Palaniappan. *Biological Signal Analysis*. 2010. ISBN: 9788776815943.

[53] P. Prandoni and M. Vetterli. *SIGNAL PROCESSING FOR COMMUNICATIONS*. 2008. ISBN: 9782940222209.

[54] A. Furdea, H. Eswaran, J. D. Wilson, H. Preissl, C. L. Lowery, and R. B. Govindan. "Magnetomyographic recording and identification of uterine contractions using Hilbert-wavelet transforms." In: *Physiological Measurement* 30.10 (2009), pp. 1051–1060. ISSN: 09673334. DOI: 10.1088/0967-3334/30/10/006.

[55] N Amrutha and V. H. Arul. "A Review on Noises in EMG Signal and its Removal." In: 7.5 (2017), pp. 23–27.

[56] C. Sousa. "Electrohysterogram Signal Component Cataloging with Spectral and Time-Frequency Methods." In: October (2015), p. 149. URL: http://hdl.handle.net/10362/16081.

[57] G Heinzel. "Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows ." In: (2002), pp. 1–84.

[58] P. D. Welch. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Aver. aging Over Short, Modified Periodograms." In: 2 (1967), pp. 70–73.

[59] I. Brain and F. With. *Imaging Brain Function With EEG*. ISBN: 9781461449836.

[60] D. G.Manolakis and V. K. Ingle. *Applied Digital Signal Processing*. 2011. ISBN: 9780521110020.

[61] P Oliveira and L Gomes. "Interpolation of signals with missing data using Principal Component Analysis." In: (2010), pp. 25–43. DOI: 10.1007/s11045-009-0086-3.

[62] M. Ermes. *Methods for the classification of biosignals applied to the detection of epileptiform waveforms and to the recognition of physical activity*. 707. 2009, pp. 1–77. ISBN: 9789513873387.

[63] I. Guyon, A. Elisseef, N. Jankowski, K. Grabczewski, G. Dreyfus, W. Duch, J. Reunanen, T. N. Lal, O. Chapelle, J. Weston, K. Torkkola, E. Tuv, M. M. Gupta, N. Homma, Z.-G. Hou, S. Gunn, A. B. Hur, G. Dror, R. M. Neal, J. Zhang, and K. Torkkola. *Supervised and unsupervised Pattern Recognition*. 205, p. 766.

[64] T.-M. Huang, V. Kecman, and I. Kopriva. *Kernel based algorithms for mining huge data sets*. 2006. ISBN: 9783540316817. DOI: http://www.zentralblatt-math.org/zmath/en/search/?an=1138.68493. arXiv: arXiv:1011.1669v3.

[65] B. Mokhlesabadifarahani and V. K. Gunjan. *EMG signals characterization in three states of contraction by fuzzy network and feature extraction*. 9789812873194. 2015. ISBN: 978-981-287-319-4. DOI: 10.1007/978-981-287-320-0.

[66] S Noelia. "Filter methods for feature selection . A comparative study." In: ().

[67] J. G. Dy and C. E. Brodley. "Feature Selection for Unsupervised Learning." In: *Journal of Machine Learning Research* 5 (2004), pp. 845–889. ISSN: 15337928. DOI: 10.1016/j.patrec.2014.11.006. URL: http://portal.acm.org/citation.cfm?id=1016787{\%}5Cnhttp://www.ece.neu.edu/fac-ece/jdy/papers/jmlr{_}tr.pdf.

[68] S. Theodoridis, K. Koutroumbas, A. Pikrakis, and D. Cavouras. *An introduction to pattern recognition - a matlab approach.* 1st ed. Academic Press, 2010, p. 231. ISBN: 9780080922751. DOI: 10.1016/B978-0-12-374486-9.00002-6.

[69] R. O.Dudat, P. E. Hart, and D. G. Stork. *Pattern Classification.*

[70] J. Irani. "Clustering Techniques and the Similarity Measures used in Clustering : A Survey." In: 134.7 (2016), pp. 9–14.